ESTIMATING MIXING TIMES:

TECHNIQUES AND APPLICATIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sharad Chandra Goel

August 2005

ESTIMATING MIXING TIMES:

TECHNIQUES AND APPLICATIONS

Sharad Chandra Goel, Ph.D.

Cornell University 2005

How many times do you have to shuffle a deck of n cards before it is close to random? $\log n$? n? n^3? Similar convergence rate questions for finite Markov chains are central to solving applied problems in diverse fields including physics, computer science and biology. This thesis investigates two general techniques for estimating mixing times for finite Markov chains: modified logarithmic Sobolev inequalities and Faber-Krahn inequalities; and analyzes the convergence behavior of a specific family of random walks: the top to bottom shuffles.

Logarithmic Sobolev inequalities are a well-studied technique for estimating convergence rates for Markov chains. In contrast to continuous state spaces, there are several distinct modified log Sobolev inequalities in the discrete setting. Here we derive modified log Sobolev inequalities for several models of random walk, including the random transposition shuffle. These results lead to tight mixing time estimates, and additionally, yield concentration inequalities.

Faber-Krahn inequalities have been used to estimate the rate of decay of the heat kernel on complete, non-compact manifolds and infinite graphs. We develop this technique in the setting of finite Markov chains, proving upper and lower L^∞ mixing time bounds via the spectral profile. This approach lets us recover previous conductance-based bounds of mixing time, and in general leads to sharper estimates of convergence rates. We apply this method to several models, including groups with moderate growth, the fractal-like Viscek graphs, and the product group $\mathbb{Z}_a \times \mathbb{Z}_b$, and obtain tight bounds on the corresponding mixing times.

A deck of n cards is shuffled by repeatedly moving the top card to one of the bottom k_n positions of the deck uniformly at random. We give upper and lower bounds on the total variation mixing time for this shuffle as k_n ranges from a constant to n. We also consider a symmetric variant of this walk which at each step either inserts the top card randomly into the bottom k_n positions or moves a random card from the bottom k_n positions to the top. For this reversible shuffle we derive L^2 mixing time bounds.

BIOGRAPHICAL SKETCH

Sharad Goel was born in Madison, Wisconsin in the summer of 1977. He moved to neighboring Illinois for college, and studied mathematics at the University of Chicago. Sharad came to Ithaca in 1999, and from Cornell he received a MS in Computer Science and a PhD in Applied Mathematics under the direction of Laurent Saloff-Coste. In the fall, he will travel to California to begin work as a Postdoctoral Scholar in Stanford's Department of Mathematics.

To Kelly

ACKNOWLEDGEMENTS

I am immensely grateful to my adviser and mentor Laurent Saloff-Coste for enthusiastically introducing me to problems, and for patiently sharing his knowledge. To Rick Durrett, I am thankful for his generous and sincere guidance. I am obliged to Jon Kleinberg for serving on my committee, and for his insightful and enjoyable class lectures. I thank Dexter Kozen for energetically proxying at my thesis defense, and for reading through my dissertation.

My friends and colleagues in the Center for Applied Mathematics have brightened my time at Cornell, and made Ithaca such an enjoyable place to work and play. I am constantly amazed by and thankful for Dolores Pendell's energy and dedication to running CAM. My family has been an enduring source of love, support and encouragement. I have benefited enormously from the kindness and generosity of Kelly Savage, my partner and friend.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

How many times do you have to shuffle a deck of n cards before it is close to random? $\log n$? n? n^3? Similar convergence rate questions for finite Markov chains are central to solving applied problems in diverse fields including physics, computer science and biology. Within this framework one can, for example, approximate the cardinality of a set Ω, and more specifically, estimate the volume of a convex body in high-dimensional Euclidean space (see e.g. [DFK91]). One can also study "typical" self-avoiding paths on a lattice, a key model for the spatial arrangement of linear polymer molecules (see e.g. [MS93]). More generally, for many applications it is useful to sample from a set of objects Ω with distribution π, where we know little about the global structure of Ω. In physics, it is common to examine properties of an "average" statistical mechanical configuration, where Ω is the set of all possible configurations and π is related to the energy of a state, for example the Gibbs distribution (see e.g. [JS93]). This approach also facilitates combinatorial optimization, where Ω is now the solution space, and π is biased toward solutions that maximize an objective function $f : \Omega \to \mathbb{R}_+$ (see e.g. [JS96]). From a Monte Carlo perspective, we can solve these problems by constructing an ergodic Markov chain with limiting distribution π, and then choosing the state of the chain at time T as the random sample. But to determine the sampling time T, simply knowing that Markov chains converge is not sufficient; mixing time estimates are essential.

Quantitative analysis of finite Markov chains benefits from both general tools to study convergence, and detailed understanding of specific examples. In line with this view, this thesis investigates two general techniques for estimating mixing time: modified logarithmic Sobolev inequalities and Faber-Krahn inequalities; and analyzes the convergence behavior of a specific family of random walks: the top to bottom shuffles.

1.1 Preliminaries

A Markov chain on a finite state space $\mathcal{X}$ can be identified with a kernel K satisfying

$$\forall x, y \quad K(x,y) \geq 0, \qquad \forall x \quad \sum_{y\in\mathcal{X}} K(x,y) = 1.$$

The iterated kernel K_n is defined by the Chapman-Kolmogorov equations

$$K_n(x,y) = \sum_{z\in\mathcal{X}} K_{n-1}(x,z)K(z,y)$$

and can be interpreted as the probability of moving from x to y in exactly n steps. We can view $K_n^x(y) = K_n(x,y)$ as a probability measure on $\mathcal{X}$. A measure π on $\mathcal{X}$ is *invariant* with respect to K if

$$\sum_{x\in\mathcal{X}} \pi(x)K(x,y) = \pi(y).$$

Throughout, we will assume that K is irreducible: For each $x, y \in \mathcal{X}$ there is an n such that $K_n(x,y) > 0$. Under this assumption K has a unique invariant probability measure π, and $\pi_* = \min_x \pi(x) > 0$. This unique probability measure is called the *stationary* or *equilibrium* distribution for the chain.

The Markov operator associated to K is defined by

$$Kf(x) = \sum_{y\in\mathcal{X}} f(y)K(x,y).$$

The iterated operator K^n then satisfies

$$K^n f(x) = \sum_{y\in\mathcal{X}} f(y)K_n(x,y).$$

The chain (K, π) is *reversible* if $K = K^*$ is a self-adjoint operator on the Hilbert space $L^2(\pi)$. This is equivalent to requiring that the *detailed balance condition* holds:

$$\forall x, y \quad \frac{K(x,y)}{\pi(y)} = \frac{K(y,x)}{\pi(x)}.$$

The kernel K describes a discrete-time chain which at each time-step moves with distribution according to K. Alternatively, we can consider the continuous-time process H_t which waits an exponential time before moving. More precisely, let $\{X_n\}$ be a Markov chain with transition probabilities given by K. Then we construct the associated continuous-time process by setting $\tilde{X}_t = X_{N_t}$ where N_t is a rate 1 Poisson process independent of $\{X_n\}$. As operators $H_t = EK_{N_t}$. Equivalently,

$$H_t = e^{-tL} \qquad L = I - K$$

where I is the identity operator. The transition kernel $H_t(x,y)$ is then given explicitly by

$$H_t(x,y) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} K_n(x,y).$$

Let $h(x,y,t) = H_t(x,y)/\pi(y)$ denote the density of $H_t^x(\cdot) = H_t(x,\cdot)$ with respect to its stationary measure π. Then for fixed $y \in \mathcal{X}$, $u(x,t) = h(x,y,t)$ is a solution to the heat equation

$$(\partial_t + L^*)u = 0 \quad u(x,t) : \mathcal{X} \times \mathbb{R}_+ \to \mathbb{R}. \tag{1.1}$$

Continuous-time, irreducible chains converge to their stationary distributions. This convergence also holds for discrete-time chains under the added assumption of aperiodicity. A state $x \in \mathcal{X}$ is *aperiodic* if $K_n(x,x) > 0$ for all sufficiently large n, and the chain is aperiodic if every state is aperiodic. For irreducible chains, aperiodicity of a single state implies aperiodicity of every state.

Proposition 1.1. *Let (K, π) be an irreducible Markov chain on a finite state space $\mathcal{X}$ with stationary distribution π. Then,*

$$\forall x, y \in \mathcal{X} \quad \lim_{t \to \infty} H_t(x, y) = \pi(y).$$

Assume further that K is aperiodic. Then,

$$\forall x, y \in \mathcal{X} \quad \lim_{n \to \infty} K_n(x, y) = \pi(y).$$

Assuming only mild conditions, Proposition 1.1 shows that Markov chains converge to stationarity. However, this classical result gives no information about the rate of convergence. *The aim of this thesis is two-fold: to present general techniques for estimating convergence to equilibrium, and to analyze the convergence behavior of specific models of random walk.*

In order to quantify a chain's distance from equilibrium we need to introduce a metric. Arguably the most natural and often used choice is the total variation distance.

Definition 1.1. Let μ and π be two probability measures on the set $\mathcal{X}$. The total variation distance is

$$\|\mu - \pi\|_{TV} = \max_{A \subset \mathcal{X}} |\mu(A) - \pi(A)|.$$

The total variation distance can be expressed in several equivalent forms as the next result shows. For details, see for example [AF].

Proposition 1.2. *Let μ and π be two probability measures on the finite set $\mathcal{X}$. Then the total variation distance satisfies*

$$\begin{aligned} \|\mu - \pi\|_{TV} &= \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \pi(x)| \\ &= \min P(V_\mu \neq V_\pi) \end{aligned}$$

where the minimum is taken over pairs of random variables (V_μ, V_π) *such that* V_μ *has distribution* μ *and* V_π *has distribution* π*.*

The second equality in Proposition 1.2 is the basis of Proposition 4.1, the well-known coupling result that is used extensively in Chapter 4 to analyze the family of top to bottom shuffles. The first equality suggests measuring distance from equilibrium in $L^p(\pi)$.

Definition 1.2. For two measures μ and ν with densities $f(x) = \mu(x)/\pi(x)$ and $g(x) = \nu(x)/\pi(x)$ with respect to the positive measure π, their $L^p(\pi)$ distance is

$$d_{\pi,p}(\mu, \nu) = \|f - g\|_{L^p(\pi)} \quad \text{for } 1 \leq p \leq \infty.$$

By Proposition 1.2, $d_{\pi,1}(\mu, \pi) = 2\|\mu - \pi\|_{TV}$. By Jensen's inequality, the function $p \mapsto d_{\pi,p}(\mu, \nu)$ is non-decreasing. In the case of reversible chains, we will often compute distance in L^2 rather than in L^1.

Next we define *mixing time*, a measure of how long it takes the chain to be close (in total variation or in L^p) to equilibrium.

Definition 1.3. The total variation mixing time $\tau(\epsilon)$ for K_n is given by

$$\tau(\epsilon) = \inf\left\{ n > 0 : \sup_x \|K_n^x - \pi\|_{TV} \leq \epsilon \right\}.$$

By convention, $\tau = \tau(1/2e)$.

Definition 1.4. The L^p mixing time $\tau_p(\epsilon)$ for K_n is given by

$$\tau_p(\epsilon) = \inf\left\{ n > 0 : \sup_{x \in \mathcal{X}} d_{p,\pi}(K_n^x, \pi) \leq \epsilon \right\}.$$

By convention $\tau_p = \tau_p(1/e)$.

Analogous definitions can be given for the continuous time chain H_t. The constant e^{-1} is chosen for convenience, but is essentially arbitrary as the next proposition shows. For details, see e.g. [AF, SC04a].

Proposition 1.3. *Fix $1 \leq p \leq \infty$, then $n \mapsto \sup_{x \in \mathcal{X}} d_{\pi,p}(K_n(x,\cdot),\pi)$ is a non-increasing sub-additive function. In particular, if*

$$\sup_{x \in \mathcal{X}} d_{\pi,p}(K_m^x, \pi) \leq \beta$$

for some fixed integer m and some $\beta \in (0,1)$ then

$$\forall n \geq m \quad \sup_{x \in \mathcal{X}} d_{\pi,p}(K_m^x, \pi) \leq \beta^{\lfloor n/m \rfloor}.$$

A particularly remarkable property exhibited by many families of finite Markov chains is the *cutoff phenomenon*: There is a small window of time during which the chains mix; before this time the walks are far from equilibrium, while after this time little additional mixing occurs. The following definitions make this precise.

Definition 1.5. Let $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)$ be an infinite family of finite chains. Then $\mathcal{F}$ presents a cutoff in total variation with critical time $\{t_n\}_1^\infty$ if $t_n \to \infty$, and $\forall \epsilon > 0$

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}_n} \|K^x_{n,(1-\epsilon)t_n} - \pi_n\|_{TV} = 1$$

and

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}_n} \|K^x_{n,(1+\epsilon)t_n} - \pi_n\|_{TV} = 0.$$

Definition 1.6. Let $\mathcal{F} = (\mathcal{X}_n, K_n, \pi_n)$ be an infinite family of finite chains. Then for $1 < p \leq \infty$, $\mathcal{F}$ presents a cutoff in L^p with critical time $\{t_n\}_1^\infty$ if $t_n \to \infty$, and $\forall \epsilon > 0$

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}_n} d_{\pi,p}(K^x_{n,(1-\epsilon)t_n}, \pi_n) = \infty$$

and

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}_n} d_{\pi,p}(K^x_{n,(1-\epsilon)t_n}, \pi_n) = 0.$$

1.2 Random Walks on Finite Groups

If the underlying state space $\mathcal{X}$ is a group, this additional structure often facilitates analyzing convergence. Let G be a finite group with probability measure q, and let $\{\eta_i\}$ be G-valued independent random variables with distribution q. The left-invariant walk on G driven by q is defined by $X_0 = e$, where e is the group identity, and

$$X_{k+1} = X_k \cdot \eta_k.$$

This corresponds to the walk on G with transition kernel $K(x, y) = q(x^{-1}y)$. Define convolution powers of q by

$$q^m(g) = q^{m-1} \star q(g) = \sum_{h\in G} q^{m-1}(h)q(h^{-1}g).$$

If supp(q) is not contained in a proper subgroup of G or in a coset of a proper normal subgroup, then

$$q^m(g) \to \frac{1}{|G|} \qquad \text{as } m \to \infty.$$

That is, for group walks the stationary measure π is always uniform.

The Markov operator Q associated to a probability measure q on G is given by $Qf = f \star q^*$ where $q^*(g) = q(g^{-1})$. The reversed random walk is driven by q^* and has as its associated operator the adjoint of Q. That is, q^* has associated Markov operator $Q^*f = f \star q$.

Furthermore,

$$\sum_{g\in G} \left|\frac{q(g)}{\pi(g)} - 1\right|^p \pi(g) = \sum_{g\in G} \left|\frac{q(g^{-1})}{\pi(g)} - 1\right|^p \pi(g).$$

Consequently, on groups $d_{\pi,p}(q) = d_{\pi,p}(q^*)$, and with respect to analyzing mixing time, we can study either the walk or its reversal.

1.3 Dirichlet Forms and Sobolev Inequalities

Ideas from functional analysis have proven to be powerful tools to investigate mixing times for finite Markov chains.

Definition 1.7. For the chain (K, π), the Dirichlet form is given by

$$\mathcal{E}(f, g) = \langle f(x), (I - K)g(x)\rangle_\pi$$

where $\langle f, g\rangle_\pi = \sum_x f(x)g(x)\pi(x)$ is the standard $L^2(\pi)$ inner product.

Proposition 1.4. *The Dirichlet form $\mathcal{E}$ satisfies:*

$$\mathcal{E}(f, f) = \langle f, (I - (K + K^*)/2)f\rangle_\pi$$

$$\mathcal{E}(f, f) = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x, y)\pi(x)$$

and

$$\frac{d}{dt}\|H_t f\|_2^2 = -2\mathcal{E}(H_t f, H_t f).$$

For reversible chains,

$$\mathcal{E}(f, g) = \frac{1}{2}\sum_{x,y}[f(x) - f(y)][g(x) - g(y)]K(x, y)\pi(x).$$

The spectral gap and logarithmic Sobolev constant are defined in terms of the Dirichlet form.

Definition 1.8. For (K, π) a Markov chain with Dirichlet form $\mathcal{E}$, the spectral gap λ_1 is defined by

$$\lambda_1 = \min\left\{\frac{\mathcal{E}(f, f)}{\mathrm{Var}_\pi(f)}; \mathrm{Var}_\pi(f) \neq 0\right\}$$

where $\mathrm{Var}_\pi(f)$ denotes the variation of f: $E(f - Ef)^2$.

While in the reversible case λ_1 is the smallest non-zero eigenvalue of $I - K$, in general λ_1 is the smallest non-zero eigenvalue of $I - \frac{1}{2}(K + K^*)$. The following proposition shows that the spectral gap controls the mixing time; details can be found in [SC96].

Proposition 1.5. *If λ_1 is the spectral gap for (K, π) then*

$$d_{\pi,2}(H_t^x, \pi) \leq \sqrt{1/\pi(x)}e^{-\lambda_1 t}.$$

It follows that

$$|H_t(x, y) - \pi(y)| \leq \sqrt{\pi(y)/\pi(x)}e^{-\lambda_1 t}.$$

Proposition 1.5 is one of the fundamental results of finite Markov chain theory. In particular, it shows that

$$\tau_2 \leq \frac{1}{2\lambda_1}\left(2 + \log \frac{1}{\pi_*}\right). \tag{1.2}$$

A common heuristic is to estimate mixing time by the *relaxation time* $1/\lambda_1$. Although Proposition 1.5 shows that this approximation is off by at most $\log(1/\pi_*)$, the log factor can be quite large: For example, walks on the symmetric group S_n have $1/\pi_* = n!$.

Logarithmic Sobolev inequalities were introduced by Gross [Gro75] to study Markov semigroups in infinite dimensional settings, and play an important role in the theory of finite Markov chains. A comprehensive overview of log Sobolev inequalities can be found in [Gro93], and [DSC96a] develops the theory for finite chains. Below we recall some fundamental results.

Definition 1.9. The entropy of a non-negative function f on $\mathcal{X}$ with respect to π is,

$$\mathrm{Ent}_\pi(f) = E\left[f \log \frac{f}{Ef}\right].$$

For an arbitrary function f, we use the notation

$$\mathcal{L}_\pi(f) = \mathrm{Ent}_\pi(f^2).$$

Observe that by Jensen's inequality applied to the convex function $\phi(t) = t\log t$, $\mathcal{L}(f) \geq 0$ and $\mathcal{L} = 0$ if and only if f is constant. The logarithmic Sobolev constant is defined analogously to the spectral gap, with $\mathrm{Var}(f)$ replaced by $\mathcal{L}(f)$.

Definition 1.10. For a Markov chain (K, π) with Dirichlet form $\mathcal{E}$, the logarithmic Sobolev constant β is defined by

$$\beta = \min\left\{\frac{\mathcal{E}(f,f)}{\mathcal{L}_\pi(f)}; \mathcal{L}_\pi(f) \neq 0\right\}.$$

From the definition, it follows that β is the largest constant c such that the logarithmic Sobolev inequality

$$c\mathcal{L}(f) \leq \mathcal{E}(f,f)$$

holds for all functions f. It is well known that $2\beta \leq \lambda_1$ (see e.g. [DSC96a]).

The following results show that the log Sobolev constant bounds entropy, which in turn bounds total variation distance. Proofs for Proposition 1.7 and Corollary 1.1 can be found in [DSC96a, SC96].

Proposition 1.6. *Let β be the log Sobolev constant for the reversible chain (K, π). Then for $f \geq 0$*

$$Ent(H_t f) \leq e^{-4\beta t} Ent(f).$$

Proof. Without loss of generality we may assume $\pi(f) = 1$. Then $\pi(H_t f) = 1$ and since $H_t^* = H_t$,

$$\begin{aligned}
\frac{d}{dt}\mathrm{Ent}(H_t f) &= \frac{d}{dt}\sum_x H_t f(x) \log H_t f(x) \pi(x) \\
&= \sum_x \left[L H_t f(x) \cdot \log H_t f(x) + L H_t f(x) \right] \pi(x) \\
&= \sum_x \left[L H_t f(x) \cdot \log H_t f(x) \right] \pi(x) \\
&= -\mathcal{E}(H_t f, \log H_t f) \\
&\leq -4\mathcal{E}((H_t f)^{1/2}, (H_t f)^{1/2}) \qquad (1.3) \\
&\leq -4\beta \mathrm{Ent}(H_t f).
\end{aligned}$$

Inequality (1.3) follows from the fact that for reversible chains

$$\forall f \geq 0, \qquad \mathcal{E}(\log f, f) \geq 4\mathcal{E}(\sqrt{f}, \sqrt{f}) \tag{1.4}$$

(see e.g. [DSC96a]). Using Gronwall's lemma, the statement is proved. $\square$

Proposition 1.7. *Let π and $\mu = h\pi$ be two probability measures on a finite set $\mathcal{X}$. Then*

$$\|\mu - \pi\|_{TV}^2 = \frac{1}{4}\|h - 1\|_{L^1(\pi)}^2 \leq \frac{1}{2} Ent_\pi(h).$$

Corollary 1.1. *If β is the log Sobolev constant for (K, π) then*

$$\|H_t^x - \pi\|_{TV}^2 \leq \frac{1}{2} \log \frac{1}{\pi(x)} \cdot e^{-4\beta t}.$$

In particular, if $\pi_* = \min_x \pi(x)$ then the mixing time satisfies

$$\tau \leq \frac{1}{4\beta}\left(3 + \log\log \frac{1}{\pi_*}\right).$$

Note that the log Sobolev constant yields a bound on mixing time with a $\log\log 1/\pi_*$ term, as opposed to the $\log 1/\pi_*$ resulting from the spectral profile bound (1.2).

1.4 Statement of Main Results

Modified Logarithmic Sobolev Inequalities. Logarithmic Sobolev inequalities were introduced by Gross in 1975 [Gro75], and can be used to estimate rates of convergence of Markov chains to their stationary distributions. While in $\mathbb{R}^n$ there are several equivalent formulations of the log Sobolev inequality, in discrete settings these formulations lead to distinct *modified* inequalities (see e.g. [BT]). Chapter 2 discusses modified log Sobolev inequalities for several models of random walk. This material also appears in [Goe04].

Theorem 1.1 gives bounds on the modified log Sobolev inequality for the random transposition chain on the symmetric group S_n, and in turn yields the correct order mixing time. Interestingly, [LY98] shows that the log Sobolev inequality yields a mixing time estimate via Corollary 1.1 that is off by a factor of $\log n$.

Like log Sobolev inequalities, modified log Sobolev inequalities can be obtained via comparison chains. Section 2.3 outlines this method, and analyzes a perturbation of the top-random transposition shuffle that cannot be realized as a walk on a group, making it difficult to study by other methods. It is well known that the Herbst argument shows that log Sobolev inequalities imply concentration inequalities (see e.g. [Led01, BT]). As an application of these results, Section 2.4 presents concentration inequalities for various models of random walk.

The recent work on modified log Sobolev inequalities [GQ03, BT] illustrates the fact that for non-diffusion Dirichlet forms, modified log Sobolev inequalities can give better results than the classical log Sobolev inequality. It is worth pointing out that the reason behind this does not seem well understood. There are, however, some drawbacks to the modified versions: First, they seem inadequate to control convergence in L^2; and second, the comparison techniques seem to be much more restricted, as discussed in Section 2.3.

Definition 1.11. For a reversible Markov chain (K, π) with Dirichlet form $\mathcal{E}$, the modified logarithmic Sobolev constant α is defined by

$$\alpha = \min \left\{ \frac{\mathcal{E}(f^2, \log f^2)}{\mathcal{L}_\pi(f)} ; \mathcal{L}_\pi(f) \neq 0 \right\}.$$

The random transposition walk on S_n is the group walk driven by the uniform measure supported on the set of all transpositions $\mathcal{C}_n = \{(i,j) | 1 \leq i < j \leq n\}$.

Theorem 1.1 (Random Transposition). *For $n \geq 2$, the random transposition walk on S_n with generating set $\mathcal{C}_n = \{(i,j) | 1 \leq i < j \leq n\}$ has modified log Sobolev constant α_n satisfying*

$$\frac{4}{n-1} \geq \alpha_n \geq \frac{1}{n-1}.$$

In particular, the total variation mixing time for the continuous-time chain satisfies

$$\tau \leq n(3 + \log \log n!).$$

The Spectral Profile. It is well known that the spectral gap of a Markov chain can be estimated in terms of conductance, facilitating isoperimetric bounds on mixing time (see [SJ89, LS88]). Observing that small sets often have large conductance, Lovász and Kannan [LK99] strengthened this result by bounding total variation mixing time for reversible chains in terms of the "average conductance" taken over sets of various sizes. Morris and Peres [MP] introduced the idea of evolving sets to analyze reversible and non-reversible chains, and found bounds on the larger L^∞ mixing time.

Chapter 3 develops Faber-Krahn inequalities in the context of finite Markov chains, side-stepping conductance to yield mixing time bounds directly in terms of the "spectral profile". Faber-Krahn inequalities were introduced by Grigor'yan and developed together with Coulhon and Pittet ([Gri94, Cou96, CGP01, BCG01]) to estimate the rate of decay of the heat kernel on manifolds and infinite graphs. Their techniques build on functional analytic methods presented, for example, in [Dav90]. We adapt this approach to the setting of finite Markov chains and derive L^∞ mixing time estimates for both reversible and non-reversible walks. The results in this chapter are joint with Ravi Montenegro and Prasad Tetali, and can also be found in [GMT].

These spectral profile bounds let us recover the previous conductance-based results, and in general lead to sharper estimates on rates of convergence to stationarity. We also show that the spectral profile can be bounded in terms of both log-Sobolev and Nash inequalities, leading to new and elementary proofs for previous mixing time results – for example, we re-derive Theorem 3.7 of Diaconis–Saloff-Coste [DSC96a] and Theorem 42 (Chapter 8) of Aldous-Fill [AF].

In terms of applications, we show optimal bounds for walks on graphs with moderate growth, walks on the fractal-like Viscek graphs, and for the group walks on the product $Z_a \times Z_b$. In the case of graphs with moderate growth, we show that the mixing time is of the order of the square of the diameter, a result originally due to Diaconis and Saloff-Coste (see [DSC94, DSC96b]). In the case of the Viscek graphs, we show that the spectral profile provides tight upper and lower bounds on mixing time, and observe that the conductance-based bounds give much weaker upper bounds.

Our main result, Theorem 1.2, bounds the L^∞ mixing time of a chain through eigenvalues of restricted Laplace operators.

Definition 1.12. For a non-empty subset $S \subset \mathcal{X}$, define

$$\lambda(S) = \inf_{f \in c_0^+(S)} \frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)}$$

where $c_0^+(S) = \{f : \mathrm{supp}(f) \subset S,\ f \geq 0\}$.

In the reversible case,

$$\lambda_0(S) \leq \lambda(S) \leq \frac{1}{1-\pi(S)}\,\lambda_0(S) \tag{1.5}$$

where λ_0 is the smallest eigenvalue of the restricted Laplacian $\Delta_S : c_0(S) \to c_0(S)$ with $c_0(S) = \{f : \mathrm{supp}(f) \subset S\}$ and

$$\Delta_S f(x) = \begin{cases} \Delta f(x) & x \in S \\ 0 & x \notin S \end{cases}$$

The kernel of $\Delta_S = I - K_S$ is given explicitly by

$$K_S(x,y) = \begin{cases} K(x,y) & x, y \in S \\ 0 & \text{otherwise} \end{cases}$$

In general, when $\pi(S) \leq 1/2$ then $\lambda(S)$ is within a factor two of the smallest eigenvalue of the symmetric operator $(\Delta_S + \Delta_S^*)/2$.

We are interested in how $\lambda(S)$ decays as the size of S increases.

Definition 1.13. Define the spectral profile $\Lambda : [\pi_*, \infty) \to \mathbb{R}$ by

$$\Lambda(r) = \inf_{\pi_* \leq \pi(S) \leq r} \lambda(S).$$

Observe that $\Lambda(r)$ is non-increasing, and $\Lambda(r) \geq \lambda_1$. By Lemma 3.2 $\Lambda(r) \approx \lambda_1$ for $r \geq 1/2$. Furthermore, by construction the walk (K, π) satisfies the *Faber-Krahn inequality*

$$\lambda(S) \geq \Lambda(\pi(S)) \quad \forall S \subset \mathcal{X}\,.$$

Theorem 1.2 (Spectral Profile). *For $\epsilon > 0$, the L^∞ mixing time $\tau_\infty(\epsilon)$ for a chain $H_t(x,y)$ satisfies*

$$\tau_\infty(\epsilon) \le \int_{4\pi_*}^{4/\epsilon} \frac{2dv}{v\Lambda(v)}\,.$$

In Section 3.2.3, we prove an analogous result for discrete-time walks. Since $\Lambda(r) \ge \lambda_1$, Theorem 1.2 shows that

$$\tau_\infty(1/e) \le \int_{4\pi_*}^{4e} \frac{2\,dv}{v\Lambda(v)} \le \frac{2}{\lambda_1}\left(1 + \log\frac{1}{\pi_*}\right).$$

But since we can expect $\Lambda(r) \gg \lambda_1$ for small r, Theorem 1.2 offers an improvement over the standard spectral gap mixing time bound (1.2).

By a discrete version of the Cheeger inequality of differential geometry,

$$\Phi_*^2(r)/2 \le \Lambda(r) \le 2\Phi_*(r)$$

where $\Phi_*(r)$ is the (truncated) conductance profile (see Section 3.2.2). Consequently, by Theorem 1.2:

Corollary 1.2. *For $\epsilon > 0$, the L^∞ mixing time $\tau_\infty(\epsilon)$ for a chain $H_t(x,y)$ satisfies*

$$\tau_\infty(\epsilon) \le \int_{4\pi_*}^{4/\epsilon} \frac{4dv}{v\Phi_*^2(v)}.$$

Theorem 13 of [MP] is a factor of two weaker than this conductance profile bound. Although Theorem 1.2 implies mixing time estimates in terms of conductance, it is reasonable to expect that for many models $\Lambda(r) \gg \Phi_*^2(r)$. In these cases, as for example the Viscek graphs of Section 3.4.4, our spectral approach leads to sharper mixing time results than does the conductance profile method.

Top to Bottom Shuffles. Chapter 4 analyzes the family of top to bottom shuffles. These shuffles are generated by moving the top card in a deck uniformly at random to any of the bottom k_n positions of the deck. For $k_n = n$ we recover the top to random walk, which exhibits cut-off at time $n \log n$ (see e.g. [Ald82, DFP92, Dia88, SC04a]). For $k_n = 2$, this is the Rudvalis shuffle, and upper and lower bounds of order $n^3 \log n$ have been shown by Hildebrand [Hil90] and Wilson [Wil03b] respectively. The following two theorems describe the mixing time of top to bottom shuffles for various ranges of k_n between these two extremes. These results are also presented in [Goe].

Theorem 1.3 (Top to Bottom Shuffles). *For the top to bottom k_n shuffles,*

1. *if $k_n \geq n - \sqrt{(n \log n)/2}$ then*

$$\tau_1 \sim n \log n.$$

That is, the walk presents a total variation cut-off at time $n \log n$.

2. *if $k_n \geq cn$ with $c \in (0,1)$ then*

$$A(c) n \log n \leq \tau_1 \leq B(c) n^2 \log^2 n$$

3. *if $k_n \leq C$ then*

$$A(C) n^3 \leq \tau_1 \leq B(C) n^3 \log n$$

4. *if $k_n = 2, 3$ then*

$$A n^3 \log n \leq \tau_1 \leq B n^3 \log n.$$

Using techniques similar to those presented in Chapter 4, Jonasson [Jon] has recently shown that top to bottom shuffles mix in $\Theta(n^3 \log n / k_n^2)$ time.

At each time step, either move the top card in a deck uniformly at random to any of the bottom k_n positions, or pick a card uniformly at random from the bottom k_n positions and move it to the top. This reversible walk is the additive symmeterization of the top to bottom shuffles.

Theorem 1.4 (Reversible Top to Bottom Shuffles). *For the reversible top to bottom k_n shuffles,*

1. *if* $k_n \geq n - C$ *then*
$$\tau_1 \leq \tau_2 \leq B(C) n \log n$$

2. *if* $k_n \leq cn$ *with* $c \in (0,1)$ *then*
$$\tau_2 \geq \tau_1 \geq A(c) n^2 \quad \text{and} \quad \tau_2 \geq \frac{A(c)n^3}{k_n^2} \log n$$

3. *for any* k_n
$$\tau_1 \leq \tau_2 \leq B n^3 \log n.$$

In particular, if $k_n \leq C$ *then*

$$A(C) n^3 \log n \leq \tau_2 \leq B n^3 \log n.$$

As k_n varies from a constant to n, these results are most satisfactory at the extremes of the range. For large k_n the walks behave like the top to random chain, mixing in $n \log n$ steps. Theorem 1.3(1) proves mixing in the strongest possible sense: cut-off at precisely $n \log n$. Let us note here that the precise L^2 cut-off time is not yet known even for the top to random shuffle $q_{n,n}$. For small k_n the walks behave like the Rudvalis shuffle, mixing in $n^3 \log n$ steps. Theorem 1.4 proves this for the reversible chain, whereas Theorem 1.3 gives complete results only for $k_n = 2, 3$.

The worst gap in these results occurs when $k_n \approx n/2$. For these "top to bottom half" shuffles, [Jon] shows a $\Theta(n \log n)$ mixing time for the non-reversible shuffle, and our results give an $\Omega(n^2)$ lower bound for the reversible shuffle. In particular, the non-reversible and reversible top to bottom half shuffles mix at different rates.

In this range, one difficulty in analyzing the reversible walk is that comparison with random transposition, one of the best understood models of random walk, can at best yield $O(n^3 \log n)$ upper bounds (see Lemma 4.10).

A variety of methods are used to prove the results of this chapter. The upper bounds for the non-reversible top to bottom shuffle are found by coupling arguments. The lower bound in Theorem 1.3(4) uses Wilson's lemma (see e.g. [SC04b, Wil03b]). For the reversible chain, we use comparison techniques for walks on finite groups to prove both upper and lower bounds (see e.g. [DSC93b]). Notably, comparison previously has been applied only to find upper bounds. It appears that this is the first application of comparison techniques to prove lower bounds.

Chapter 2

Modified Log Sobolev Inequalities

Introduced in 1975 [Gro75], logarithmic Sobolev inequalities can be used to estimate rates of convergence of Markov chains to their stationary distributions. While in $\mathbb{R}^n$ there are several equivalent formulations of the log Sobolev inequality, in discrete settings these formulations lead to distinct inequalities (see e.g. [BT]). One such modification, considered in [Wu00, GQ03, Rob01], is the topic of this chapter. Much of the work presented here can also be found in [Goe04].

In Section 2.1, we introduce the modified log Sobolev inequality α. As a first example, we discuss estimates of α for the walk on the 2-point space (see also [BT]). Section 2.2 proves the main results of this chapter: modified logarithmic Sobolev inequalities for several models of random walk, including the random transposition shuffle and the top-random transposition shuffle on the symmetric group, and the shuffle generated by 3-cycles on the alternating group. As an application of these results we derive sharp bounds on rates of convergence. Previously, convergence results for these models had been obtained by Fourier analysis [DS81, Dia92, Rou00]. In this section we also show that a generic r-regular graph has modified log Sobolev constant much smaller than its spectral gap. After completing this work, it came to our attention that Gao and Quastel [GQ03] had derived the modified log Sobolev inequality for the random transposition model.

Like log Sobolev inequalities, modified log Sobolev inequalities can be obtained via comparison chains. Section 2.3 outlines this method, and analyzes a perturbation of the top-random transposition shuffle that cannot be realized as a walk on a group, making it difficult to study by other methods. It is well known that the Herbst argument shows that log Sobolev inequalities imply concentration inequalities (see e.g. [Led01, BT]). As an application of our results, in Section 2.4 we present concentration inequalities for the models of random walk mentioned above.

The recent work on modified log Sobolev inequalities [GQ03, BT] illustrates the fact that for non-diffusion Dirichlet forms, modified log Sobolev inequalities can give better results than the classical log Sobolev inequality. It is worth pointing out that the reason behind this does not seem well understood. There are, however, some drawbacks to the modified versions: First, they seem inadequate to control convergence in L^2; and second, the comparison techniques seem to be much more restricted, as discussed in Section 2.3.

2.1 Background

Proposition 1.6 shows that the log Sobolev constant controls entropy. That proof requires two key inequalities: For reversible chains and $f \geq 0$,

$$\mathcal{E}(f, \log f) \geq 4\mathcal{E}(\sqrt{f}, \sqrt{f})$$

and

$$\beta\mathcal{L}(f) \leq \mathcal{E}(f, f).$$

The second inequality comes from the definition of the log Sobolev constant. The modified log Sobolev constant is motivated by the fact that to control entropy we only need a bound on $\mathcal{E}(f^2, \log f^2)$ (see e.g. [Rob01, BT, Wu00]).

Definition 2.1. For a Markov chain (K, π) with Dirichlet form $\mathcal{E}$, the modified logarithmic Sobolev constant α is defined by

$$\alpha = \min\left\{\frac{\mathcal{E}(f^2, \log f^2)}{\mathcal{L}_\pi(f)}; \mathcal{L}_\pi(f) \neq 0\right\}.$$

Modified log Sobolev inequalities have recently been studied in several settings: In [Wu00], modified log Sobolev inequalities were found for Poisson measures on $\mathbb{N}$; and [Rob01] derives them for birth and death process on $\mathbb{Z}$. For a discussion of several different discrete modifications of the log Sobolev inequality, see e.g. [BT, ABC+00, BL98, AL00, BL00]

We have the following well known result relating the log Sobolev constant, the modified log Sobolev constant and the spectral gap.

Proposition 2.1. *For a reversible chain (K, π) the log Sobolev constant β, the modified log Sobolev constant α and the spectral gap λ_1 satisfy*

$$4\beta \le \alpha \le 2\lambda_1.$$

Proof. The first inequality follow from (1.4). The proof of the second inequality is analogous to corresponding inequality for the standard log Sobolev constant. Fix g and set $f = 1 + \epsilon g$ for ϵ small enough so that $|\epsilon g| < 1$.

$$\begin{aligned} f^2 \log f^2 &= 2(1 + 2\epsilon g + \epsilon^2 g^2)\left(\epsilon g - \frac{\epsilon^2 g^2}{2} + O(\epsilon^3)\right) \\ &= 2\epsilon g + 3\epsilon^2 g^2 + O(\epsilon^3) \end{aligned}$$

and

$$\begin{aligned} f^2 \log \|f\|_2^2 &= (1 + 2\epsilon g + \epsilon^2 g^2)(2\epsilon Eg - \epsilon^2 \|g\|_2^2 - 2\epsilon^2 (Eg)^2 + O(\epsilon^3)) \\ &= 2\epsilon Eg + 4\epsilon^2 g Eg + \epsilon^2 \|g\|_2^2 - 2\epsilon^2 (Eg)^2 + O(\epsilon^3). \end{aligned}$$

So,

$$f^2 \log \frac{f^2}{\|f\|_2^2} = 2\epsilon(g - Eg) + \epsilon^2 \left(3g^2 - \|g\|_2^2 - 4gEg + 2(Eg)^2\right) + O(\epsilon^3)$$

and

$$\begin{aligned} \mathcal{L}(f) &= 2\epsilon^2 \left(\|g\|_2^2 - (Eg)^2\right) + O(\epsilon^3) \\ &= 2\epsilon^2 \mathrm{Var}(g) + O(\epsilon^3). \end{aligned}$$

Now,

$$\begin{aligned}
\mathcal{E}(f^2, \log f^2) &= 2\mathcal{E}\left(1 + 2\epsilon g + \epsilon^2 g^2, \epsilon g - \frac{\epsilon^2 g^2}{2} + O(\epsilon^3)\right) \\
&= 4\epsilon^2 \mathcal{E}(g,g) + 2\mathcal{E}\left(1, \epsilon g - \frac{\epsilon^2 g^2}{2} + O(\epsilon^3)\right) + \\
&\quad 2\mathcal{E}\left(\epsilon^2 g^2, \epsilon g - \frac{\epsilon^2 g^2}{2} + O(\epsilon^3)\right) \\
&= 4\epsilon^2 \mathcal{E}(g,g) + 2\mathcal{E}\left(\epsilon^2 g^2, \epsilon g - \frac{\epsilon^2 g^2}{2} + O(\epsilon^3)\right) \\
&= 4\epsilon^2 \mathcal{E}(g,g) + O(\epsilon^3).
\end{aligned}$$

Consequently,

$$\alpha \leq \frac{\mathcal{E}(f^2, \log f^2)}{\mathcal{L}(f)} = 2\frac{\epsilon^2 \mathcal{E}(g,g) + O(\epsilon^3)}{\epsilon^2 \mathrm{Var}(g) + O(\epsilon^3)}.$$

Taking the limit as $\epsilon \to 0$, we have

$$\alpha \leq 2\frac{\mathcal{E}(g,g)}{\mathrm{Var}(g)}.$$

Minimizing over g, we have the result.

$\square$

From the definition we see that α is the largest constant c such that the modified log Sobolev inequality

$$c\mathcal{L}(f) \leq \mathcal{E}(f^2, \log f^2)$$

holds for all functions f. Consequently, as in the case of the log Sobolev inequality, α controls entropy, and in turn mixing time.

Proposition 2.2. *Let α be the modified log Sobolev constant for the chain (K, π). Then for $f \geq 0$*

$$Ent(H_t^* f) \leq e^{-\alpha t} Ent(f).$$

Proof. Without loss of generality we may assume $\pi(f) = 1$. Then $\pi(H_t^* f) = 1$ and

$$\begin{aligned}
\frac{d}{dt}\mathrm{Ent}(H_t^* f) &= \frac{d}{dt}\sum_x H_t^* f(x) \log H_t^* f(x)\pi(x) \\
&= \sum_x \left[L^* H_t^* f(x) \cdot \log H_t^* f(x) + L^* H_t^* f(x)\right]\pi(x) \\
&= \sum_x \left[L^* H_t^* f(x) \cdot \log H_t^* f(x)\right]\pi(x) \\
&= -\mathcal{E}(H_t^* f, \log H_t^* f) \\
&\leq -\alpha \mathrm{Ent}(H_t^* f).
\end{aligned}$$

By Gronwall's lemma we have the result.

$\square$

Corollary 2.1. *If α is the modified log Sobolev constant for the reversible chain (K, π) then*

$$\|H_t^x - \pi\|_{TV}^2 \leq \frac{1}{2}\log\frac{1}{\pi(x)} \cdot e^{-\alpha t}.$$

In particular, if $\pi_* = \min_x \pi(x)$ then the mixing time satisfies

$$\tau \leq \frac{1}{\alpha}\left(3 + \log\log\frac{1}{\pi_*}\right).$$

Proof. Let δ_x denote the Dirac mass function. Then

$$\begin{aligned}
H_t^*\left(\frac{\delta_x}{\pi(x)}\right)(y) &= \frac{H_t^*(y,x)}{\pi(x)} \\
&= \frac{H_t(x,y)}{\pi(y)}.
\end{aligned}$$

So,

$$\begin{aligned}
\|H_t^x - \pi\|_{TV}^2 &= \frac{1}{4}\|\frac{H_t^x(\cdot)}{\pi(\cdot)} - 1\|_{L^1(\pi)}^2 \\
&\leq \frac{1}{2}\mathrm{Ent}_\pi\left(\frac{H_t^x(\cdot)}{\pi(\cdot)}\right) \text{ by Proposition 1.7} \\
&= \frac{1}{2}\mathrm{Ent}_\pi\left(H_t^* \frac{\delta_x}{\pi(x)}\right) \\
&\leq \frac{1}{2}e^{-\alpha t}\mathrm{Ent}_\pi\left(\frac{\delta_x}{\pi(x)}\right) \text{ by Proposition 2.2} \\
&= \frac{1}{2}\log\frac{1}{\pi(x)} \cdot e^{-\alpha t}.
\end{aligned}$$

□

The modified log Sobolev constant is a phenomenon of the discrete state space. Let $d\mu(x) = w(x)dx$ be a probability measure on $\mathbb{R}^n$ with a smooth strictly positive density w. Then the continuous analog of the discrete Dirichlet form is,

$$\mathcal{E}(f,g) = \int_{\mathbb{R}^n} \nabla f(x) \cdot \nabla g(x) d\mu(x)$$

where ∇ is the usual gradient. In this setting, since we have a chain rule,

$$\mathcal{E}(f^2, \log f^2) = 4\mathcal{E}(f,f).$$

On discrete state spaces, (1.4) shows that we have only inequality, suggesting that in this setting α and β may differ. However, given that they are indistinguishable on $\mathbb{R}^n$, it is surprising that we do in fact find examples where $\alpha \gg \beta$.

The modified log Sobolev and log Sobolev inequalities share several properties, two of which we state here. The first shows that the modified log Sobolev inequality behaves well under products, and the second shows that solutions to the modified log Sobolev inequality satisfy a certain difference equation. Lemma 2.1 and Theorem 2.1 are analogous to statements for the log Sobolev inequality given in [DSC96a, SC96].

Lemma 2.1. *Let (K_i, π_i), $i = 1, \dots d$, be reversible Markov chains on finite sets $\mathcal{X}_i$ with modified log Sobolev constants α_i. Fix $\mu = (\mu_i)_i^d$ such that $\mu_i > 0$ and $\sum \mu_i = 1$. Then the product chain (K, π) on $\mathcal{X} = \prod_i^d \mathcal{X}_i$ with kernel*

$$K(x,y) = \sum_{i=1}^{d} \mu_i \delta(x_1, y_1) \dots \delta(x_{i-1}, y_{i-1}) K_i(x_i, y_i) \delta(x_{i+1}, y_{i+1}) \dots \delta(x_d, y_d)$$

(where $\delta(x,x) = 1$ and $\delta(x,y) = 0$ for $x \neq y$) and stationary measure $\pi = \bigotimes \pi_i$ satisfies

$$\alpha = \min_i \mu_i \alpha_i.$$

Proof. Let $\mathcal{E}_i$ denote the Dirichlet form associated to K_i. Then the product K chain has Dirichlet form

$$\mathcal{E}(f,g) \;=\; \sum_{i=1}^{d} \mu_i \left(\sum_{\substack{x_j \in \mathcal{X}_j \\ j \neq i}} \mathcal{E}_i(f,g)(x^i) \pi^i(x^i) \right)$$

where x^i is the sequence $(x_1, \dots, x_{i-1}, x_{i+1}, \dots x_d)$, $\pi^i = \bigotimes_{l \neq i} \pi_l$, and $\mathcal{E}(f,g)(x^i)$ is to be interpreted as $\mathcal{E}_i$ acting on f and g considered as functions of their i^{th} coordinates with the remaining coordinates given by x^i. By induction, it is only necessary to prove the statement when $d = 2$. Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \Re$ be a nonnegative function and let

$$\begin{aligned} F(x_2)^2 &= \left(\sum_{x_1} f(x_1, x_2)^2 \pi_1(x_1) \right) \\ &= E_{\pi_1} f(\cdot, x_2)^2. \end{aligned}$$

Then, $\|F\|_{2,\pi_2} = \|f\|_{2,\pi}$. And

$$\begin{aligned} \mathcal{L}(f) &= \sum_{x_1, x_2} f(x_1, x_2)^2 \log \frac{f(x_1, x_2)^2}{\|f\|_{2,\pi}^2} \pi(x_1, x_2) \\ &= \sum_{x_2} F(x_2)^2 \log \frac{F(x_2)^2}{\|F\|_{2,\pi_2}^2} \pi_2(x_2) + \sum_{x_1, x_2} f(x_1, x_2)^2 \log \frac{f(x_1, x_2)^2}{F(x_2)^2} \pi(x_1, x_2) \\ &\leq \frac{1}{\alpha_2} \mathcal{E}_2(F^2, \log F^2) + \frac{1}{\alpha_1} \sum_{x_2} \mathcal{E}_1 \left(f(\cdot, x_2)^2, \log f(\cdot, x_2)^2 \right) \pi_2(x_2). \end{aligned}$$

Now, since $g(x,y) = (x-y)(\log x - \log y)$ is convex for $x, y > 0$, by Jensen's inequality we have

$$\begin{aligned}
&[F(x_2)^2 - F(y_2)^2][\log F(x_2)^2 - \log F(y_2)^2] \\
&\quad = [E_{\pi_1} f(\cdot, x_2)^2 - E_{\pi_1} f(\cdot, y_2)^2][\log E_{\pi_1} f(\cdot, x_2)^2 - \log E_{\pi_1} f(\cdot, y_2)^2] \\
&\quad \leq E_{\pi_1}[f(\cdot, x_2)^2 - f(\cdot, y_2)^2][\log f(\cdot, x_2)^2 - \log f(\cdot, y_2)^2].
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\mathcal{E}_2(F^2, \log F^2) \\
&\quad = \frac{1}{2} E_{\pi_2}\left[\sum_{y_2} [F(x_2)^2 - F(y_2)^2][\log F(x_2)^2 - \log F(y_2)^2] K_2(x_2, y_2)\right] \\
&\quad \leq \frac{1}{2} E_{\pi_1} E_{\pi_2}\left[\sum_{y_2} [f(\cdot, x_2)^2 - f(\cdot, y_2)^2][\log f(\cdot, x_2)^2 - \log f(\cdot, y_2)^2] K_2(x_2, y_2)\right] \\
&\quad = \sum_{x_1} \mathcal{E}_2(f(x_1, \cdot)^2, \log f(x_1, \cdot)^2) \pi_1(x_1).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{L}(f) &\leq \frac{\mu_2}{\mu_2 \alpha_2} \sum_{x_1} \mathcal{E}_2\left(f(\cdot, x_2)^2, \log f(\cdot, x_2)^2\right) \pi_1(x_1) + \\
&\quad \frac{\mu_1}{\mu_1 \alpha_1} \sum_{x_2} \mathcal{E}_1\left(f(\cdot, x_2)^2, \log f(\cdot, x_2)^2\right) \pi_2(x_2) \\
&\leq \max_i \frac{1}{\mu_i \alpha_i} \mathcal{E}(f^2, \log f^2).
\end{aligned}$$

So, $\alpha \geq \min_i \mu_i \alpha_i$. Testing on functions that only depend on one of the two variables yields the result. □

Theorem 2.1. *Let (K, π) be a reversible Markov chain with modified log Sobolev constant α and spectral gap λ_1. Then either $\alpha = 2\lambda_1$ or there exists a positive, non-constant function u which is a solution of*

$$u^2 \log u^2 - u^2 \log \|u\|_2^2 - \frac{1}{\alpha} u^2 (I - K) \log u^2 - \frac{1}{\alpha}(I - K) u^2 = 0 \tag{2.1}$$

and satisfies

$$\alpha\mathcal{L}(u) = \mathcal{E}(u^2, \log u^2).$$

In particular, if K is irreducible, then $\alpha > 0$.

Proof. When looking for a minimizer of $\Gamma(f) = \mathcal{E}(f^2, \log f^2)/\mathcal{L}(f)$, we need only consider non-negative functions f with $\pi(f) = 1$ (since $\mathcal{E}(f^2, \log f^2)$ and $\mathcal{L}(f)$ are homogeneous of degree 2). Note that $A = \{f | \pi(f) = 1\}$ is a closed, bounded, and hence compact, subset of R^n. Furthermore, the discontinuities of Γ in A occur at $f \equiv 1$ and $f(x_i) = 0$ for some $1 \leq i \leq n$. Therefore, either there exists a non-constant, positive minimizer $u \in A$ of Γ or the minimum α is attained as a limit towards one of the discontinuity points of Γ in A.

First consider a sequence $f_m \to f$ with $f_m \in A$ and $f(x_i) = 0$ for some $1 \leq i \leq n$. Since f is non-constant, $\mathcal{L}(f) > 0$. Now consider one term

$$\left[f_m^2(x_i) - f_m^2(x_j)\right]\left[\log f_m^2(x_i) - \log f_m^2(x_j)\right]$$

of the sum in the definition of $\mathcal{E}(f^2, \log f^2)$ where $f_m(x_j) \to c > 0$. Then the product goes to $+\infty$ and since all of the terms in the sum are non-negative, $\Gamma(f_m) \to +\infty$.

Next consider the case where $f_m \to 1$. Define, $g_m = f_m - 1$. Observe that $\mathcal{E}(f, g) \leq 2\|f\|_\infty \|g\|_\infty$. So from the proof of Proposition 2.1 we have that,

$$\begin{aligned}
\Gamma(f_m) &= \Gamma(1 + g_m) \\
&= 2\frac{\mathcal{E}(g_m, g_m) + O(\|g_m\|_\infty^3)}{\mathrm{Var}(g_m) + O(\|g_m\|_\infty^3)} \\
&\geq \left(2\lambda + \frac{O(\|g_m\|_\infty^3)}{\mathrm{Var}(g_m)}\right) / \left(1 + \frac{O(\|g_m\|_\infty^3)}{\mathrm{Var}(g_m)}\right).
\end{aligned}$$

Since we have a finite state space, and $Eg_m = 0$,

$$\mathrm{Var}(g_m) = \|g_m\|_2^2 \approx \|g_m\|_\infty^2.$$

So,

$$\frac{O(\|g_m\|_\infty^3)}{\mathrm{Var}(g_m)} \to 0,$$

and by Proposition 2.1, $\alpha = 2\lambda_1$. So we have shown that either $\alpha = 2\lambda_1$ or there exists a positive, non-constant minimizer u of Γ. Now fix $u > 0$, a minimizer of Γ, and any function g. Then note that,

$$\frac{d}{d\tau}\Gamma(u+\tau g)(0) = 0.$$

And,

$$\begin{aligned}
\mathcal{E}\left((u+\tau g)^2, \log(u+\tau g)^2\right) &= 2\mathcal{E}\left(u^2 + 2ug\tau + \tau^2 g^2, \log\left(1+\tau\frac{g}{u}\right) + \log u\right) \\
&= 2\mathcal{E}\left(u^2 + 2ug\tau + \tau^2 g^2, \tau\frac{g}{u} + \log u + 0(\tau^2)\right) \\
&= 2\mathcal{E}\left(u^2, \log u\right) + 4\tau\mathcal{E}\left(ug, \log u\right) + 2\tau\mathcal{E}\left(u^2, \frac{g}{u}\right) + 0(\tau^2).
\end{aligned}$$

Since (K, π) is reversible, i.e K is self-adjoint with respect to $L^2(\pi)$, we get

$$\begin{aligned}
\frac{d}{d\tau}\mathcal{E}\left((u+\tau g)^2, \log(u+\tau g)^2\right)(0) &= 4\mathcal{E}\left(ug, \log u\right) + 2\mathcal{E}\left(u^2, \frac{g}{u}\right) \\
&= 4\left(\log u, (I-K)ug\right)_\pi + 2\mathcal{E}\left(\frac{g}{u}, (I-K)u^2\right)_\pi \\
&= 4\left((I-K)\log u, ug\right)_\pi + 2\mathcal{E}\left(\frac{g}{u}, (I-K)u^2\right)_\pi \\
&= 2\left(u(I-K)\log u^2 + \frac{1}{u}(I-K)u^2, g\right)_\pi.
\end{aligned}$$

Furthermore, since

$$\begin{aligned}
\|u+\tau g\|_2^2 &= \|u\|_2^2 + 2\tau\sum_x ug\pi(x) + \tau^2\|fg\|_2^2 \\
&= \|u\|_2^2\left(1 + 2\tau\frac{\sum_x ug\pi(x)}{\|u\|_2^2} + \frac{\tau^2\|fg\|_2^2}{\|u\|_2^2}\right)
\end{aligned}$$

we have that

$$\begin{aligned}
&(u+\tau g)^2 \log \frac{(u+\tau g)^2}{\|u+\tau g\|_2^2} \\
&= (u^2+2\tau ug+g^2)\left(\log u^2 + 2\tau\frac{g}{u} - \log\|u\|_2^2 - 2\tau\frac{\sum_x ug\pi(x)}{\|u\|_2^2} + O(\tau^2)\right) \\
&= u^2\log\frac{u^2}{\|u\|_2^2} + 2\tau\left(ug - \frac{u^2\sum_x ug\pi(x)}{\|u\|_2^2} + ug\log u^2 - ug\log\|u\|_2^2\right) + O(\tau^2).
\end{aligned}$$

So,

$$\mathcal{L}(u+\tau g) = \mathcal{L}(u) + 2\tau\left(u\log u^2 - u\log\|u\|_2^2, g\right) + O(\tau^2).$$

Consequently,

$$\frac{d}{d\tau}\mathcal{L}(u+\tau g)(0) = 2\left(u\log u^2 - u\log\|u\|_2^2, g\right),$$

Finally,

$$\begin{aligned}
0 &= \frac{d}{d\tau}\Gamma(u+\tau g)(0) \\
&= \frac{\frac{d}{d\tau}\mathcal{E}\left((u+\tau g)^2, \log(u+\tau g)^2\right)(0)}{\mathcal{L}(u)} - \frac{\mathcal{E}(u^2, \log\frac{u^2}{\|u\|_2^2})\frac{d}{d\tau}\mathcal{L}(u+\tau g)(0)}{\mathcal{L}^2(u)}
\end{aligned}$$

and,

$$\left(u\log u^2 - u\log\|u\|_2^2 - \frac{1}{\alpha}u(I-K)\log u^2 - \frac{1}{\alpha u}(I-K)u^2, g\right) = 0.$$

Since the above holds for all g, multiplying through by u gives the result. $\square$

The symmetric walk on the 2-point space $\mathcal{X} = \{x_1, x_2\}$ is perhaps the simplest of all Markov chains. The kernel K is given by

$$K(x_1,x_2) = K(x_2,x_1) = 1$$

$$K(x_1,x_1) = K(x_2,x_2) = 0$$

and the stationary measure π is uniform. In [Gro75], it is shown that the log Sobolev inequality for this walk satisfies $\beta = 1$. A trivial computation shows that the spectral gap $\lambda_1 = 2$. Consequently, by Proposition 2.1, the modified log Sobolev constant satisfies $\alpha = 4$. By Lemma 2.1 and the fact that both the spectral gap and the log Sobolev constant are also well-behaved under products (see e.g. [SC96]), the walk on the n-dimensional hypercube has constants satisfying $4\beta = \alpha = 2\lambda = \frac{4}{n}$.

A generalization of the walk on the 2-point space is the complete walk on n-points addressed in the following lemma.

Lemma 2.2. *Consider the Markov chain on the n point space $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ with uniform kernel $U(x_i, x_j) = \frac{1}{n-1}$ for $x_i \neq x_j$ and $U(x_i, x_i) = 0$. For $n \geq 2$, the modified log Sobolev constant α_n satisfies*

$$\frac{n}{n-1} \leq \alpha_n \leq \left(1 + \frac{4}{\log(n+1)}\right) \frac{n}{n-1}.$$

Proof. Since the chain has the uniform stationary distribution $\pi(x_i) = \frac{1}{n}$, we have

$$\begin{aligned}
\mathcal{E}(f^2, \log f^2) &= \frac{1}{2n(n-1)} \sum_{i,j=1}^{n} [f^2(x_i) - f^2(x_j)][\log f^2(x_i) - \log f^2(x_j)] \\
&= \frac{n}{n-1} \left(E[f^2 \log f^2] - Ef^2 \cdot E \log f^2\right). \\
&= \frac{n}{n-1} \left(E\left[f^2 \log \frac{f^2}{Ef^2}\right] - Ef^2 \cdot E \log \frac{f^2}{Ef^2}\right).
\end{aligned}$$

By Jensen's inequality $E \log \frac{f^2}{Ef^2} \leq 0$ and the lower bound is established. For the upper bound, take f with $f^2(x_1) = n + 1$ and $f^2(x_i) = 1$ for $2 \leq x_i \leq n$. $\square$

[DSC96a] proves that for the complete walk on $\mathcal{X}_n$ the log Sobolev constant satisfies

$$\beta = \frac{1 - 2/n}{\log(n-1)},$$

showing that for this example, $\alpha \gg \beta$.

An alternative generalization of the symmetric walk on the 2-point space is the asymmetric walk of Corollary 2.2.

Corollary 2.2. *Consider the Markov chain on the two point space* $\mathcal{X}_2 = \{x_1, x_2\}$ *with kernel* $K(x_i, x_1) = \rho$ *and* $K(x_i, x_2) = 1 - \rho$ *with* $i = 1, 2$ *and* $0 < \rho \leq 1/2$. *Then the modified log Sobolev constant satisfies* $1 \leq \alpha \leq 2$.

Proof. (K, π) is a reversible chain with stationary distribution $\pi(x_1) = \rho$, $\pi(x_2) = 1 - \rho$. First we establish the lower bound, observing that it is sufficient to restrict our attention to functions with $Ef = 1$. Consider rational ρ and write $\rho = \frac{p}{q}$ for integer p, q. Since we can identify functions f on $\mathcal{X}_2$ with functions $\tilde{f}$ on $\mathcal{X}_q = \{x_1, \ldots x_q\}$ that are constant on the subsets $\{x_1, \ldots, x_p\}$ and $\{x_{p+1}, \ldots, x_q\}$, Lemma 2.2 shows that

$$\begin{aligned}\rho f^2(x_1) \log f^2(x_1) + (1-\rho) f^2(x_2) \log f^2(x_2) \leq \\ \rho(1-\rho)[f^2(x_1) - f^2(x_2)][\log f^2(x_1) - \log f^2(x_2)].\end{aligned}$$

The result for irrational ρ follows by holding f fixed and taking the limit as $\rho_n \to \rho$ for rational $\{\rho_n\}$. The upper bound follows from the fact that the spectral gap $\lambda = 1$, and Proposition 2.1. $\square$

For the asymmetric walk, the log Sobolev constant was calculated in [DSC96a] (and also independently in [HY]) and shown to satisfy

$$\beta = \frac{1 - 2\rho}{\log[(1-\rho)/\rho]},$$

again exemplifying the difference between α and β. For a further discussion of the asymmetric walk, see [BT].

2.2 Modified Logarithmic Sobolev Inequalities

In this section we derive modified log Sobolev inequalities for some models of random walk, including the random transposition shuffle and the top-random transposition shuffle on the symmetric group, and the walk generated by 3-cylces on the alternating group. These results are used to deduce sharp bounds on mixing times. We also show that a generic r-regular graph has modified log Sobolev constant much smaller than its spectral gap.

2.2.1 Random Transposition and Related Walks

The random transposition walk on the symmetric group S_n is a shuffle on a deck of n cards where we uniformly at random select and swap pairs of cards. The convergence behavior of this model was studied in detail in [DS81] using the representation theory of the symmetric group. The log Sobolev constant for this walk was determined in [LY98] to satisfy $\beta^{-1} \asymp n \log n$. Surprisingly, the log Sobolev constant is inadequate to sharply bound the mixing time via Corollary 1.1. Using the martingale method of [LY98], Theorem 2.2 bounds the modified log Sobolev constant for walks including and related to random transposition. In contrast to the log Sobolev constant, our estimate of the modified log Sobolev constant for random transposition is sufficiently strong to yield the correct mixing time. After this work was completed, it came to our attention that Gao and Quastel [GQ03] had proven Theorem 2.2 for the case of random transposition.

Let $G_n \subset S_n$ be subgroups of the symmetric group, generated by the symmetric sets $\mathcal{C}_n \subset G_n$. Then we have associated random walks given by the kernel,

$$K_n(\tau,\tau') = \begin{cases} \frac{1}{|\mathcal{C}_n|} & \tau' = \tau\cdot\sigma \text{ for some } \sigma \in \mathcal{C}_n \\ 0 & \text{otherwise} \end{cases}.$$

That is, K_n is the kernel of a group walk driven by the uniform measure supported on C_n. The stationary distribution π_n is uniform on G_n and the Dirichlet form for (K_n, π_n) is given explicitly by

$$\mathcal{E}_n(f,g) = \frac{1}{2|\mathcal{C}_n|} E\left[\sum_{\tau'\in\mathcal{C}_n} [f(\tau) - f(\tau\cdot\tau')][g(\tau) - g(\tau\cdot\tau')]\right].$$

For $\tau \in S_n$, we let τ_i denote the particle in position i, and so $\tau\cdot\sigma$ denotes the configuration after we permute the positions according to σ. If this Markov chain has enough symmetry, Theorem 2.2 gives a bound on the modified log Sobolev constant α.

Definition 2.2. A sequence of groups $G_n \subset S_n$ with symmetric generating sets $\mathcal{C}_n$ is called self-similar if:

1. For $1 \le s \le n$, there exist isomorphisms $g_s^{n-1} : G_{n-1} \to \{\sigma \in G_n | \sigma_s = s\}$, with $g_s^{n-1}(\mathcal{C}_{n-1}) = \{\sigma \in \mathcal{C}_n | s \notin \operatorname{supp}(\sigma)\}$
2. G_n acts transitively on the set $\{1, \ldots, n\}$
3. There exists k, such that for all n and $\sigma \in \mathcal{C}_n$, $|\operatorname{supp}(\sigma)| = k$, where $\operatorname{supp}(\sigma) = \{i|\sigma_i \neq i\}$.

Definition 2.2 encompasses a collection of random walks including random transposition on S_n and the walk generated by 3-cycles on the alternating group A_n. More generally, consider a sequence of random walks generated by conjugacy classes of S_n. Recall, that for $n \neq 4$, a non-trivial conjugacy class $\mathcal{C}_n$ generates either the alternating group A_n or S_n. For a permutation $\tau \in S_n$, let

$c(\tau) = (c_1, \ldots, c_n)$ denote the cycle structure of τ. That is, c_i is the number of cycles of length i in the disjoint cycle decomposition of τ. Then, two permutations are conjugate if and only if their cycle structure is the same. Now, for a conjugacy class $\mathcal{C}_{n_0}$ of S_{n_0} (respectively A_{n_0}) with corresponding cycle structure $c^{n_0} = (c_1^{n_0}, \ldots, c_{n_0}^{n_0})$, define a sequence of conjugacy classes $\mathcal{C}_n$ for $n > n_0$ corresponding to the cycle structures $c^n = (c_1^n, c_2^{n_0}, \ldots, c_{n_0}^{n_0}, 0, \ldots, 0)$ where $c_1^n = n - \sum_{i=2}^{n_0} i c_i^{n_0}$. Then this sequence of walks is self-similar.

For $1 \leq s \leq n$, let $\sigma_s : S_n \to \{x_1, \ldots, x_n\}$ be the random variable that takes $\tau \to \tau_s$. The idea behind the proof of Theorem 2.2 is to first condition on each σ_s. Then we break up $\mathcal{L}(f)$ into two parts: The first we bound by the Dirichlet form on S_{n-1} where we have fixed the s^{th} position to hold particle σ_s. The second we bound by looking at the complete walk described in Lemma 2.2 with stationary measure corresponding to the distribution of σ_s (i.e. the uniform distribution on $\{1, \ldots, n\}$). By averaging over s, we can pass from the Dirichlet forms on S_{n-1} to S_n. This then gives us a recurrence relation between the modified log Sobolev constants, yielding the result.

Theorem 2.2. *Let $\mathcal{C}_n \subset G_n$ be self-similar for $n \geq n_0$, and consider the sequence of walks generated as above. Then, if a_n denotes the reciprocal of the modified log Sobolev constant for these chains,*

$$a_n \leq a_{n_0} + (n - n_0).$$

Proof. To begin we fix a function $f : S_n \to \Re$, with $f > 0$. By homogeniety, it is sufficient to establish the modified log Sobolev inequality for f with $\pi(f^2) = 1$. Define,

$$
\begin{aligned}
f_s(x) &= E[f^2(\cdot)|\sigma_s = x]^{1/2} \\
f_s(\tau|x) &= \frac{f(\tau)\delta_x(\tau_s)}{f_s(x)}
\end{aligned}
$$

where δ_x is the dirac point mass at x. Let

$$
\begin{aligned}
I_{1,s} &= E\Bigg\{ f_s^2(\sigma_s) E\Bigg[\sum_{\substack{\tau' \in \mathcal{C}_n \\ supp(\tau') \not\ni s}} [f_s^2(\tau|\sigma_s) - f_s^2(\tau \cdot \tau'|\sigma_s)] \\
&\qquad \times [\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau \cdot \tau'|\sigma_s)] \Bigg| \sigma_s \Bigg] \Bigg\} \\
&= E\Bigg\{ f_s^2(\sigma_s) E\Bigg[\sum_{\substack{\tau' \in \mathcal{C}_n \\ supp(\tau') \not\ni s}} \left[\frac{f^2(\tau)}{f_s^2(\sigma_s)} - \frac{f^2(\tau \cdot \tau')}{f_s^2(\sigma_s)} \right] \\
&\qquad \times \left[\log \frac{f^2(\tau)}{f_s^2(\sigma_s)} - \log \frac{f^2(\tau \cdot \tau')}{f_s^2(\sigma_s)} \right] \Bigg| \sigma_s \Bigg] \Bigg\} \\
&= E\left(E\left[\sum_{\substack{\tau' \in \mathcal{C}_n \\ supp(\tau') \not\ni s}} \left[f^2(\tau) - f^2(\tau \cdot \tau') \right] \left[\log f^2(\tau) - \log f^2(\tau \cdot \tau') \right] \Bigg| \sigma_s \right] \right) \\
&= E\left[\sum_{\substack{\tau' \in \mathcal{C}_n \\ supp(\tau') \not\ni s}} \left[f^2(\tau) - f^2(\tau \cdot \tau') \right] \left[\log f^2(\tau) - \log f^2(\tau \cdot \tau') \right] \right].
\end{aligned}
$$

And define,

$$
\begin{aligned}
I_{2,s} &= E\left[f_s^2(\sigma_s) \log f_s^2(\sigma_s) \right] \\
I_i &= \frac{1}{n} \sum_{s=1}^{n} I_{i,s} \qquad\qquad i = 1, 2.
\end{aligned}
$$

To express $\mathcal{L}(f)$ in terms of the above definitions note that

$$
\begin{aligned}
f_s^2(\sigma_s) E\left[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s) \middle| \sigma_s \right] &= f_s^2(\sigma_s) E\left[\frac{f^2(\tau)}{f_s^2(\sigma_s)} \log \frac{f^2(\tau)}{f_s^2(\sigma_s)} \middle| \sigma_s \right] \\
&= E\left[f^2(\tau) \log f^2(\tau) \middle| \sigma_s \right] - f_s^2(\sigma_s) \log f_s^2(\sigma_s).
\end{aligned}
$$

Taking expectations,

$$
\mathcal{L}(f) = E\left\{ f_s^2(\sigma_s) E\left[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s) \middle| \sigma_s \right] \right\} + I_{2,s}. \tag{2.2}
$$

Since $f_s^2(\tau|\sigma_s) = 0$ for $\tau_s \neq \sigma_s$, we can naturally consider $f_s(\cdot|\sigma_s)$ as a function on S_{n-1} (where we fix position s to hold particle σ_s). Specifically, let $h_{s\to\sigma_s} \in G_n$ be such that $h_{s\to\sigma_s}(s) = \sigma_s$, and define $\tilde{f}_s^2$ on G_{n-1} by

$$\tilde{f}_s^2(\tau) = f_s^2\left(h_{s\to\sigma_s} g_s^{n-1}(\tau)|\sigma_s\right).$$

Since $E_{n-1}\tilde{f}_s^2(\cdot) = 1$,

$$\begin{aligned}
E\left[f_s^2(\tau|\sigma_s)\log f_s^2(\tau|\sigma_s)\middle|\,\sigma_s\right] &= E\left[\tilde{f}_s^2\log\tilde{f}_s^2\right] \\
&\leq \frac{a_{n-1}}{2|\mathcal{C}_{n-1}|}E\left[\sum_{\tau'\in\mathcal{C}_{n-1}}[\tilde{f}_s^2(\tau)-\tilde{f}_s^2(\tau\cdot\tau')][\log\tilde{f}_s^2(\tau)-\log\tilde{f}_s^2(\tau\cdot\tau')]\right] \\
&= \frac{a_{n-1}}{2|\mathcal{C}_{n-1}|}E\Bigg[\sum_{\substack{\tau'\in\mathcal{C}_n\\ supp(\tau')\not\ni s}}[f_s^2(\tau|\sigma_s)-f_s^2(\tau\cdot\tau'|\sigma_s)] \\
&\qquad \times[\log f_s^2(\tau|\sigma_s)-\log f_s^2(\tau\cdot\tau'|\sigma_s)]\Bigg|\sigma_s\Bigg].
\end{aligned}$$

Applying this to (2.2) gives,

$$\mathcal{L}(f) \leq \frac{a_{n-1}}{2|\mathcal{C}_{n-1}|}I_{1,s} + I_{2,s}.$$

Averaging over s, we have

$$\mathcal{L}(f) \leq \frac{a_{n-1}}{2|\mathcal{C}_{n-1}|}I_1 + I_2. \tag{2.3}$$

Let $k = \text{supp}(\sigma)$ for $\sigma \in \mathcal{C}_n$. Then note that each term

$$[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau')]$$

appears in I_1 exactly $n-k$ times. So we have,

$$\begin{aligned}
I_1 &= \frac{n-k}{n}E\left[\sum_{\tau'\in\mathcal{C}_n}[f^2(\tau)-f^2(\tau\cdot\tau')][\log f^2(\tau)-\log f^2(\tau\cdot\tau')]\right] \\
&= \frac{2(n-k)|\mathcal{C}_n|}{n}\mathcal{E}(f^2,\log f^2).
\end{aligned}$$

Note that

$$\begin{aligned}
|\mathcal{C}_n|(n-k) &= \sum_{\sigma\in\mathcal{C}_n}\sum_{i=1}^{n} 1_{\{i|i\notin supp(\sigma)\}} \\
&= \sum_{i=1}^{n}\sum_{\sigma\in\mathcal{C}_n} 1_{\{i|i\notin supp(\sigma)\}} \\
&= n|\mathcal{C}_{n-1}|.
\end{aligned}$$

Substituting this into (2.3), we have

$$\mathcal{L}(f) \leq a_{n-1}\mathcal{E}(f^2, \log f^2) + I_2. \tag{2.4}$$

To bound I_2 we consider the Markov chain on state space $\{x_1,\ldots,x_n\}$ with uniform kernel $K(x_i,x_j) = \frac{1}{n-1}$ for $i \neq j$. First note that since $|\{\tau|\tau_s = i\}| = |\{h_{s\to i}\{\tau|\tau_s = s\}| = |G_{n-1}|$ for all i, σ_s is uniformly distributed on $\{x_1,\ldots,x_n\}$. Furthermore,

$$\begin{aligned}
E[f_\cdot^2(x_m)] &= \frac{1}{n}\sum_{s=1}^{n} E[f^2|\sigma_s = x_m] \\
&= \sum_{s=1}^{n} E[f^2; \sigma_s = x_m] \\
&= 1.
\end{aligned}$$

Consequently,

$$\begin{aligned}
I_2 &= \frac{1}{n}\sum_{s=1}^{n} E\left[f_s^2(\cdot)\log f_s^2(\cdot)\right] \\
&= \frac{1}{n}\sum_{m=1}^{n} \mathcal{L}_U\left(f_\cdot(x_m)\right) \\
&\leq \frac{n-1}{n^2}\sum_{m=1}^{n} \mathcal{E}\left(f_\cdot^2(x_m), \log f_\cdot^2(x_m)\right) \qquad \text{by Lemma 2.2} \\
&= \frac{1}{2n^3}\sum_{m=1}^{n}\sum_{i\neq j}[f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)].
\end{aligned}$$

Now, for $\tau' \in \mathcal{C}_n$ such that $\tau'_i = j$,

$$\begin{aligned}
&[f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)] \\
&= \big[E[f^2(\tau)|\sigma_i = x_m] - E[f^2(\tau)|\sigma_j = x_m]\big] \\
&\quad \times \big[\log E[f^2(\tau)|\sigma_i = x_m] - \log E[f^2(\tau)|\sigma_j = x_m]\big] \\
&= \big[E[f^2(\tau\cdot\tau')|\sigma_j = x_m] - E[f^2(\tau)|\sigma_j = x_m]\big] \\
&\quad \times \big[\log E[f^2(\tau\cdot\tau')|\sigma_j = x_m] - \log E[f^2(\tau)|\sigma_j = x_m]\big] \\
&\leq E\left[[f^2(\tau) - f^2(\tau\cdot\tau')][\log f^2(\tau) - \log f^2(\tau\cdot\tau')]\,\middle|\, \sigma_j = x_m\right] \\
&= nE\left[f^2(\tau) - f^2(\tau\cdot\tau')][\log f^2(\tau) - \log f^2(\tau\cdot\tau'); \sigma_j = x_m\right].
\end{aligned}$$

Above we have used Jensen's inequality with $g(x, y) = (x-y)(\log x - \log y)$ (which is convex for $x, y > 0$). Let $\mathcal{C}_{i\to j} \subset \mathcal{C}_n$ consist of those τ' with $\tau'_i = j$. Then averaging over $\tau' \in \mathcal{C}_{i\to j}$, we get

$$\begin{aligned}
&[f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)] \\
&\leq \frac{n}{|\mathcal{C}_{i\to j}|} \sum_{\tau' \in \mathcal{C}_{i\to j}} E\left[f^2(\tau) - f^2(\tau\cdot\tau')][\log f^2(\tau) - \log f^2(\tau\cdot\tau'); \sigma_j = x_m\right].
\end{aligned}$$

And,

$$\begin{aligned}
n(n-1)|\mathcal{C}_{i\to j}| &= \sum_{i\neq j} \sum_{\sigma\in\mathcal{C}_n} 1_{\{\sigma_i = j\}} \\
&= \sum_{\sigma\in\mathcal{C}_n} \sum_{i\neq j} 1_{\{\sigma_i = j\}} \\
&= |\mathcal{C}_n| k
\end{aligned}$$

yields,

$$\begin{aligned}
I_2 &\leq \frac{n-1}{2nk|\mathcal{C}_n|} \sum_{m=1}^{n} \sum_{i \neq j} \sum_{\tau' \in \mathcal{C}_{i \to j}} E\Big[[f^2(\tau) - f^2(\tau \cdot \tau')] \\
&\quad \times [\log f^2(\tau) - \log f^2(\tau \cdot \tau')]; \sigma_j = x_m\Big] \\
&= \frac{n-1}{2nk|\mathcal{C}_n|} \sum_{i \neq j} \sum_{\tau' \in \mathcal{C}_{i \to j}} E\left[f^2(\tau) - f^2(\tau \cdot \tau')][\log f^2(\tau) - \log f^2(\tau \cdot \tau')\right] \\
&= \frac{n-1}{2n|\mathcal{C}_n|} \sum_{\tau' \in \mathcal{C}_n} E\left[f^2(\tau) - f^2(\tau \cdot \tau')][\log f^2(\tau) - \log f^2(\tau \cdot \tau')\right]. \\
&= \frac{n-1}{n} \mathcal{E}(f^2, \log f^2).
\end{aligned}$$

Using (2.4), the result follows from the recurrence,

$$a_n \leq a_{n-1} + \frac{n-1}{n}.$$

□

As a corollary to Theorem 2.2 we have Theorem 1.1, one of the main results of this thesis.

Theorem 1.1. For $n \geq 2$, the random transposition walk on S_n with generating set $\mathcal{C}_n = \{(i, j) | 1 \leq i < j \leq n\}$ has modified log Sobolev constant α_n satisfying

$$\frac{4}{n-1} \geq \alpha_n \geq \frac{1}{n-1}.$$

In particular, the total variation mixing time for the continuous-time chain satisfies

$$\tau \leq n(3 + \log \log n!).$$

Proof. Since the walk on S_2 is the symmetric walk on the 2-point space, the discussion of that example in Section 2.1 shows that $a_2 = \frac{1}{4}$, yielding the lower bound. This chain is studied in detail in [Dia88], where it is shown that the spectral gap satisfies $\lambda_1 = \frac{2}{n-1}$. The upper bound is then a consequence of Proposition 2.1. The mixing time follows from Corollary 2.1. □

Remark 2.1. In [SC94] it is shown that for $t = \frac{n-1}{2}(\log(n-1) - c)$ the random transposition shuffle satisfies

$$\|H_t^x - \pi\|_{TV} \geq 1 - 8e^{-2c} - 4e^{-c} - \frac{4\log(n-1)}{n-1} e^{\frac{2\log n}{n}}$$

and results in [DS81] show that this bound is sharp. Consequently, the mixing time bound of Theorem 1.1 is within a factor of 2 of the critical time.

Corollary 2.3. *For the random walk on A_n generated by 3-cycles, the modified log Sobolev constant satisfies*

$$\frac{6}{n-1} \geq \alpha \geq \frac{1}{n-2}.$$

In particular, the mixing time satisfies $\tau \leq n(3 + \log\log n!)$.

Remark 2.2. In [Rou99, Rou00] it is shown that the above walk has cutoff with critical time $t_n = \frac{n}{3}\log n$.

Proof. The walk on A_3 is the uniform walk on the 3-point space. Consequently, by Lemma 2.2, $a_3 \leq 1$, yielding the lower bound. The upper bound follows from results in [Rou99, Rou00] that $\lambda = \frac{3}{n-1}$, and Proposition 2.1. $\square$

Example 2.1. Informally, the Bernoulli-Laplace (BL) model is a random transposition walk on S_n with n distinct sites and $1 \leq r \leq n-1$ identical particles, with each site occupied by at most one particle. The state space $C_{n,r}$ is the set of r-subsets of $\{1, \ldots, n\}$, and accordingly is of order $\binom{n}{r}$. For $\eta \in C_{n,r}$ let η_i denote the number of particles in site i, so η_i is either 0 or 1. Let η^{ij} denote the configuration in which we have swapped the particles in positions i and j. Then the kernel for this chain is given by

$$K(\eta, \eta') = \begin{cases} \frac{1}{r(n-r)} & \eta' = \eta^{ij} \qquad \eta_i = (1-\eta_j) = 1 \\ 0 & \text{otherwise} \end{cases}$$

The log Sobolev constant for this walk was found in [LY98] to satisfy

$$\beta_{n,r} \asymp \frac{n}{r(n-r)} \log \frac{n^2}{r(n-r)},$$

and in [GQ03] the authors used the method of Theorem 2.2 to directly show that modified log Sobolev constant for BL satisfies $\alpha_{n,r} \asymp \frac{n}{r(n-r)}$. We can find this same bound on the modified log Sobolev constant by relying on our analysis of the random transposition walk.

To analyze this chain, map functions f on $C_{n,r}$ to functions $\tilde{f}$ on S_n by letting $\tilde{f}(\sigma) = f(\{\sigma_1, \ldots, \sigma_r\})$. For $\eta \in C_{n,r}$ let $\overline{\eta} \in S_n$ be any permutation such that $\eta = \{\overline{\eta}_1, \ldots, \overline{\eta}_r\}$. Note that there are $r!(n-r)!$ such permutations and that $f(\eta^{ij}) = \tilde{f}(\overline{\eta}^{ij})$. Therefore,

$$\begin{aligned}
\mathcal{E}(f,g) &= \frac{1}{2r(n-r)} E_\pi \left[\sum_{\substack{i,j \\ \eta_i = (1-\eta_j)=1}} [f(\eta) - f(\eta^{ij})][g(\eta) - g(\eta^{ij})] \right] \\
&= \frac{1}{4r(n-r)} \frac{r!(n-r)!}{n!} \left[\sum_{i,j} [f(\eta) - f(\eta^{ij})][g(\eta) - g(\eta^{ij})] \right] \\
&= \frac{1}{4r(n-r)} \frac{1}{n!} \left[\sum_{i,j} [\tilde{f}(\sigma) - \tilde{f}(\sigma^{ij})][\tilde{g}(\sigma) - \tilde{g}(\sigma^{ij})] \right] \\
&= \frac{n(n-1)}{2r(n-r)} \mathcal{E}'(\tilde{f}, \tilde{g})
\end{aligned}$$

where $\mathcal{E}'(f,g)$ is the Dirichlet form associated with the random transposition model (K', π') of Theorem 1.1. Furthermore, since $\mathcal{L}(f) = \mathcal{L}'(\tilde{f})$, $\alpha_{n,r} \geq \frac{n(n-1)}{2r(n-r)} \alpha'_n$. By Theorem 1.1 and the fact that the spectral gap for Bernoulli-Laplace is given by $\lambda_{n,r} = \frac{n}{r(n-r)}$ [DS87], the modified log Sobolev constant for the BL model satisfies

$$\frac{2n}{r(n-r)} \geq \alpha_{n,r} \geq \frac{n}{2r(n-r)}.$$

By Corollary 2.1, the mixing time $\tau_{n,r}$ for the BL model satisfies

$$\tau_{n,r} \leq \frac{2r(n-r)}{n} \left(3 + \log \log \binom{n}{r} \right).$$

2.2.2 Top-Random Transposition Walk

A walk similar to those considered in Section 2.2.1 is the complete (k,l)-bipartite shuffle. We can visualize this walk on a deck of cards by first splitting the deck into two pieces–of size k and of size l–and then uniformly at random swapping pairs of cards between the piles. In the case $k=1$, we have the top-random transposition shuffle. That is, at each step we swap the top card and another chosen uniformly at random.

Using the same notation as above, for $\tau \in S_n$, we let τ_i denote the particle in position i, and let τ^{ij} denote the configuration after we swap the particles in positions i and j. The kernel for the (k,l)-complete bipartite shuffle is given by,

$$K(\tau,\tau') = \begin{cases} \frac{1}{kl} & \tau' = \tau^{ij} \text{ for any } 1 \le i \le k < j \le k+l \\ 0 & \text{otherwise} \end{cases}$$

The stationary distribution π is uniform and the Dirichlet form for (K,π) is given explicitly by

$$\mathcal{E}(f,g) = \frac{1}{2kl} E\left[\sum_{1\le i\le k<j\le k+l} [f(\tau)-f(\tau^{ij})][g(\tau)-g(\tau^{ij})]\right].$$

As above, before computing the modified log Sobolev constant for the walk on S_n, we restrict our attention to the movement of one particle. In this case we have the walk on the complete (k,l)-bipartite graph: Our state space is the $k+l$ point space $\{x_1,\ldots,x_k,y_1,\ldots y_l\}$; the kernel is given by

$$\begin{aligned}
\tilde{K}(x_i, y_j) &= \frac{1}{k+l} && 1 \le i \le k,\ 1 \le j \le l \\
\tilde{K}(y_j, x_i) &= \frac{1}{k+l} && 1 \le i \le k,\ 1 \le j \le l \\
\tilde{K}(x_i, x_i) &= \frac{k}{l+k} && 1 \le i \le k \\
\tilde{K}(y_j, y_j) &= \frac{l}{l+k} && 1 \le j \le l
\end{aligned}$$

and zero otherwise. Then $(\tilde{K}, \pi)$ is reversible with respect to the uniform stationary measure π.

Lemma 2.3. *For the random walk on the complete (k,l)-bipartite graph with $l \ge 2$, the modified log Sobolev constant satisfies $\alpha \le \frac{2k}{k+l}$. In the case of the star, i.e. the complete $(1,l)$-bipartite graph, we have the lower bound $\alpha \ge \frac{1}{l+1}$.*

Proof. By explicitly computing the eigenvalues of $\tilde{K}$ we find the spectral gap $\lambda = \frac{k}{k+l}$. The upper bound for α then follows from Proposition 2.1.

To lower bound α for the star observe that

$$\begin{aligned}
\mathcal{E}(f^2, \log f^2) &= \frac{1}{(l+1)^2}\left[\sum_i [f^2(x_1) - f^2(y_i)][\log f^2(x_1) - \log f^2(y_i)]\right] \\
&= \frac{1}{l+1} E\left[[f^2(x_1) - f^2(\cdot)][\log f^2(x_1) - \log f^2(\cdot)]\right].
\end{aligned}$$

By homogeneity, we only need to show the modified log Sobolev inequality for functions f with $f^2(x_1) = 1$. And in this case, the above simplifies to

$$\mathcal{E}(f^2, \log f^2) = \frac{1}{l+1}\left[-E\log f^2 + Ef^2 \log f^2\right].$$

By Jensen's inequality,

$$E \log f^2 \le \log Ef^2 \le Ef^2 \cdot \log Ef^2.$$

Since $\mathcal{L}(f) = Ef^2 \log f^2 - Ef^2 \cdot \log Ef^2$,

$$\mathcal{L}(f) \leq Ef^2 \log f^2 - E \log f^2$$

and the lower bound is established. □

The proof of the following theorem is analogous to the proof of Theorem 2.2, the primary difference being that here we bound I_2 using the Markov chain on the star described in Lemma 2.3.

Theorem 2.3. *For $l \geq 2$, let $a_{k,l}$ denote the reciprocal of the modified log Sobolev constant for the complete (k,l)-bipartite walk on S_{k+l}, and let $\tilde{a}_{k,l}$ denote the reciprocal of the modified log Sobolev constant for the complete (k,l)-bipartite walk on $\{x_1, \ldots, x_k, y_1, \ldots, y_l\}$. Then*

$$a_{k,l} \leq a_{k,l-1} + \frac{2k}{k+l} \tilde{a}_{k,l}.$$

Proof. For $1 \leq s \leq k+l$, let $\sigma_s : S_n \to \{x_1, \ldots, x_n\}$ be the random variable that takes $\tau \to \tau_s$. To begin we fix a function $f : S_n \to \Re$, with $f > 0$ By homogeneity, it is sufficient to establish the modified log Sobolev inequality for f with $\pi(f^2) = 1$. Let

$$\begin{aligned} f_s(x) &= E[f^2(\cdot)|\sigma_s = x]^{1/2} \\ f_s(\tau|x) &= \frac{f(\tau)\delta_x(\tau_s)}{f_s(x)} \end{aligned}$$

And for $k < s \leq k+l$, define

$$\begin{aligned}
I_{1,s} &= E\Bigg\{ f_s^2(\sigma_s) E\Bigg[\sum_{\substack{1\leq i\leq k<j\leq k+l \\ j\neq s}} [f_s^2(\tau|\sigma_s) - f_s^2(\tau^{ij}|\sigma_s)] \\
&\qquad \times [\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau^{ij}|\sigma_s)] \Bigg| \sigma_s \Bigg] \Bigg\} \\
&= E\Bigg[\sum_{\substack{1\leq i\leq k<j\leq k+l \\ j\neq s}} \left[f^2(\tau) - f^2(\tau^{ij})\right] \left[\log f^2(\tau) - \log f^2(\tau^{ij})\right] \Bigg] \\
I_{2,s} &= E\left[f_s^2(\sigma_s) \log f_s^2(\sigma_s) \right] \\
I_i &= \frac{1}{l} \sum_{s=k+1}^{k+l} I_{i,s} \qquad i = 1, 2.
\end{aligned}$$

As before,

$$\mathcal{L}(f) = E\left\{ f_s^2(\sigma_s) E\left[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s) \middle| \sigma_s \right] \right\} + I_{2,s},$$

and, considering $f_s(\cdot|\sigma_s)$ as a function on S_{n-1},

$$\mathcal{L}_{k,l-1}(f_s(\cdot|\sigma_s)) \leq a_{k,l-1} \mathcal{E}_{k,l-1}(f_s^2(\cdot|\sigma_s), \log f_s^2(\cdot|\sigma_s)).$$

Consequently,

$$\begin{aligned}
E\left[f_s^2(\tau|\sigma_s) \log f_s^2(\tau|\sigma_s) \middle| \sigma_s \right] &\leq \frac{a_{k,l-1}}{2k(l-1)} E\Bigg[\sum_{\substack{1\leq i\leq k<j\leq k+l \\ j\neq s}} [f_s^2(\tau|\sigma_s) - f_s^2(\tau^{ij}|\sigma_s)] \\
&\qquad \times [\log f_s^2(\tau|\sigma_s) - \log f_s^2(\tau^{ij}|\sigma_s)] \Bigg| \sigma_s \Bigg].
\end{aligned}$$

and

$$\mathcal{L}(f) \leq \frac{a_{k,l-1}}{2k(l-1)} I_{1,s} + I_{2,s}.$$

Averaging over s, we have

$$\mathcal{L}(f) \leq \frac{a_{k,l-1}}{2k(l-1)} I_1 + I_2. \tag{2.5}$$

Since each term $[f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij})]$ appears in I_1 exactly $l-1$ times, we have

$$\begin{aligned} I_1 &= \frac{l-1}{l} E\left[\sum_{1 \le i \le k < j \le k+l} [f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij})] \right] \\ &= 2k(l-1)\mathcal{E}(f^2, \log f^2). \end{aligned}$$

Substituting this into (2.5), we have

$$\mathcal{L}(f) \le a_{k,l-1}\mathcal{E}(f^2, \log f^2) + I_2. \tag{2.6}$$

To bound I_2 we consider the Markov chain on the state spaces $\{x_1, \ldots, x_{k+l}\}$ with kernel given by the complete (k,l)-bipartite graph. Since σ_s is uniformly distributed on $\{1, \ldots, k+l\}$,

$$\begin{aligned} I_2 &= \frac{1}{l} \sum_{s=k+1}^{k+l} E\left[f_s^2(\cdot) \log f_s^2(\cdot) \right] \\ &\le \frac{1}{l} \sum_{s=1}^{k+l} E\left[f_s^2(\cdot) \log f_s^2(\cdot) \right] \qquad \text{since entropy is non-negative} \\ &= \frac{1}{l} \sum_{m=1}^{k+l} \mathcal{L}_U\left(f_\cdot(x_m) \right) \\ &\le \frac{\tilde{a}_{k,l}}{l} \sum_{m=1}^{k+l} \mathcal{E}\left(f_\cdot^2(x_m), \log f_\cdot^2(x_m) \right) \\ &= \frac{\tilde{a}_{k,l}}{l(k+l)^2} \sum_{m=1}^{k+l} \sum_{1 \le i \le k < j \le k+l} [f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)]. \end{aligned}$$

Since,

$$\begin{aligned} & [f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)] \\ & \le \; (k+l)E\left[f^2(\tau) - f^2(\tau^{ij})][\log f^2(\tau) - \log f^2(\tau^{ij}); \sigma_j = x_m \right], \end{aligned}$$

we have,

$$\begin{aligned} I_2 &\le \frac{\tilde{a}_{k,l}}{l(k+l)} \sum_{1 \le i \le k < j \le k+l} [f_i^2(x_m) - f_j^2(x_m)][\log f_i^2(x_m) - \log f_j^2(x_m)]. \\ &= \frac{2k}{k+l} \tilde{a}_{k,l} \mathcal{E}(f^2, \log f^2) \end{aligned}$$

and the corresponding recurrence,

$$a_{k,l} \leq a_{k,l-1} + \frac{2k}{k+l}\tilde{a}_{k,l}.$$

$\square$

Corollary 2.4. *For the top-random transposition walk on S_n, i.e. the complete $(1, n-1)$-bipartite shuffle, the modified log Sobolev constant satisfies*

$$\frac{2}{n-1} \geq \alpha \geq \frac{1}{2(n-1)}.$$

In particular, the mixing time satisfies $\tau \leq 2(n-1)[3 + \log\log n!]$.

Remark 2.3. [Dia92] outlines a proof that the top-random transposition walk exhibits total variation cutoff at critical time $t_n = n \log n$.

Proof. By Lemma 2.3, $\tilde{a}_{1,l} \leq l + 1$ and the recurrence reduces to $a_{1,l} \leq a_{1,l-1} + 2$. Since the top-random transposition walk on S_2 is the symmetric walk on the 2-point space, $a_{1,1} = \frac{1}{4}$, yielding the lower bound. The upper bound follows from Proposition 2.1 and the fact that $\lambda_1 = \frac{1}{n-1}$ (see [FOW85]). $\square$

2.2.3 Random Regular Graphs

In the examples that we have examined thus far, the modified log Sobolev constant α is of approximately the same magnitude as the spectral gap λ_1. This is not, however, always the case.

For fixed $r > 3$, [Bol80] introduced a model for random r-regular graphs on n vertices. For this model, [Alo87] shows that as n tends to infinity, a random r-regular graph $\mathcal{G}$ has spectral gap $\lambda_1(\mathcal{G}) \geq \epsilon(r) > 0$ with probability $1 - o(1)$. For this model, [DSC96a] shows that a generic r-regular graph has log Sobolev constant

$\beta \ll \lambda$. The following lemma shows that for this family of random graphs we also have $\alpha \ll \lambda$.

Lemma 2.4. *Let $G = (\mathcal{X}, E)$ be a connected r-regular graph with $|\mathcal{X}| \geq 8r$ and let (K, π) be the canonical walk on G with uniform stationary distribution π. Then the modified log Sobolev constant satisfies*

$$\alpha \leq 4 \log r \frac{2 + \log\log|\mathcal{X}|}{\log[|\mathcal{X}|/8]}.$$

In particular, for fixed r

$$\alpha \overset{|\mathcal{X}|\to\infty}{\longrightarrow} 0.$$

Proof. For $x, y \in \mathcal{X}$, let $d(x, y)$ be the natural graph distance. Let $A_x^n = \{y \in \mathcal{X} \mid d(x, y) > n\}$. Then $K_x^n(A_x^n) = 0$, and

$$\begin{aligned} \pi(A_x^n) &\geq 1 - \frac{1 + r + r^2 + \cdots + r^n}{|\mathcal{X}|} \\ &\geq 1 - \frac{2r^n}{|\mathcal{X}|} \end{aligned}$$

since by assumption we must have $r > 1$. Let $\mathrm{Po}(\lambda)$ denote a Poisson random variable with mean λ. Then,

$$H_t^x(A_x^n) \leq \mathrm{Prob}(\mathrm{Po}(t) \geq n) \leq \frac{t + t^2}{n^2}.$$

Furthermore for $n_0 = \frac{\log(|\mathcal{X}|/8)}{\log r}$ and $t_0 = \frac{n_0}{4}$, $\pi(A_x^{n_0}) \geq \frac{3}{4}$ and $H_{t_0}^x(A_x^{n_0}) \leq \frac{1}{4} + \frac{1}{16}$. Since,

$$\|H_x^{t_0} - \pi\|_{TV} = \max_{A \subset \mathcal{X}} |H_x^{t_0}(A) - \pi(A)| \geq \frac{7}{16}$$

the mixing time satisfies $\tau > t_0$. But from Lemma 2.1,

$$\tau \leq \frac{1}{\alpha}(\log\log|\mathcal{X}| + 2)$$

and the result follows. $\square$

2.3 Comparison Techniques

A perturbed chain often has log Sobolev and modified log Sobolev constants similar to the original. Lemma 2.5 illustrates this phenomenon; the proof is similar to the log Sobolev case presented in [DSC96a] and is omitted here.

Lemma 2.5. *Let (K, π) and (K', π') be two finite, reversible Markov chains defined on $\mathcal{X}$ and $\mathcal{X}'$ respectively with modified log Sobolev constants α and α'. Assume there exists a map*

$$l^2(\mathcal{X}, \pi) \to l^2(\mathcal{X}', \pi') \ : \ f \to \tilde{f}$$

and constants $A, B, a > 0$ such that for all $f \in l^2(\mathcal{X}, \pi)$

$$\mathcal{E}'(\tilde{f}^2, \log \tilde{f}^2) \leq A\mathcal{E}(f^2, \log f^2) \text{ and } a\mathcal{L}_\pi(f) \leq \mathcal{L}_{\pi'}(\tilde{f}) + B\mathcal{E}(f^2, \log f^2),$$

then

$$\frac{a\alpha'}{A + B\alpha'} \leq \alpha.$$

In the case $\mathcal{X} = \mathcal{X}'$, $\mathcal{E}'(f^2, \log f^2) \leq A\mathcal{E}(f^2, \log f^2)$ and $a\pi \leq \pi'$, we have

$$\frac{a\alpha'}{A} \leq \alpha.$$

[SC96] shows that for any two finite irreducible Markov chains K' and K on the same state space, there exists a constant A such that for all functions f, $\mathcal{E}'(f, f) \leq A\mathcal{E}(f, f)$. Consequently, the analog of Lemma 2.5 proven in [DSC96a] shows that we can always use the log Sobolev constant of one chain to estimate the constant for any other chain on the same space–although in practice this estimate may be quite bad. However, there does not in general exist a constant A such that for all f, $\mathcal{E}'(f^2, \log f^2) \leq A\mathcal{E}(f^2, \log f^2)$. Given the numerous similarities between the log Sobolev and modified log Sobolev constant, this fact is quite surprising.

Consider the three-point space $\mathcal{X} = \{x_1, x_2, x_3\}$. Let K' be the complete graph on $\mathcal{X}$, and let K be the line graph with holding probability $\frac{1}{2}$ at the endpoints. Then both chains have uniform stationary distribution. Let

$$\Delta_{x_i x_j}(f) = [f^2(x_i) - f^2(x_j)][\log f^2(x_i) - \log f^2(x_j)].$$

Then,

$$\frac{\mathcal{E}'(f^2, \log f^2)}{\mathcal{E}(f^2, \log f^2)} = 1 + \frac{\Delta_{x_1 x_3}}{\Delta_{x_1 x_2} + \Delta_{x_2 x_3}}.$$

Fix $A > 1$ and for $b > A$ let $f^2(x_1) = 1$, $f^2(x_2) = \frac{b}{A}$, and $f^2(x_3) = b$. Then

$$\begin{aligned} \frac{\Delta_{x_1 x_3}}{\Delta_{x_1 x_2} + \Delta_{x_2 x_3}} &= \frac{(b-1)\log b}{\left(\frac{b}{A} - 1\right)\log \frac{b}{A} + \frac{A-1}{A} b \log A} \\ &\geq \frac{(b-1)\log b}{\frac{b}{A}\log b + b \log A} \\ &\overset{b\to\infty}{\longrightarrow} A. \end{aligned}$$

So for every A, there exists an f with $\mathcal{E}'(f^2, \log f^2) > A\mathcal{E}(f^2, \log f^2)$. While this shows that we cannot always compare chains, several interesting examples are amenable to comparison.

Example 2.2. Recall our informal description of the top-random transposition walk on the permutation group S_n: Uniformly at random pick a position i from 2 to n, and then swap the top card and the card in position i. More formally, this is the group walk on S_n with generating set $\{(1, i) \mid 2 \leq i \leq n\}$. Consider the following variant of this walk: Uniformly at random pick a position i from 2 to n; if either the top card or the card in position i is the 'Ace of Spades', do nothing with probability $\frac{1}{2}$ and swap the cards with probability $\frac{1}{2}$; otherwise, swap the cards as usual. Like the top-random transposition walk, this modified walk is reversible with respect to the uniform stationary distribution. However, unlike the former walk, the latter cannot be realized as a group walk. While intuitively this small perturbation should not dramatically affect mixing time, the comparison techniques of [DSC93b] to obtain precise results rely crucially on group structure.

The kernel of the Heavy Ace walk is given explicitly by

$$\tau_1 \neq 1 \qquad K(\tau,\tau') = \begin{cases} \frac{1}{(n-1)} & \tau' = \tau^{1i} \quad 1 < i \leq n,\ \tau_i \neq 1 \\ \frac{1}{2(n-1)} & \tau' = \tau^{1i} \quad 1 < i \leq n,\ \tau_i = 1 \\ \frac{1}{2(n-1)} & \tau' = \tau \quad \tau_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\tau_1 = 1 \qquad K(\tau,\tau') = \begin{cases} \frac{1}{2(n-1)} & \tau' = \tau^{1i} \quad 1 < i \leq n \\ \frac{1}{2} & \tau' = \tau \quad \tau_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

Letting $\mathcal{E}'$ be the Dirichlet form of the top-random transposition walk, we see that $\mathcal{E}'(f^2, \log f^2) \leq 2\mathcal{E}(f^2, \log f^2)$. By Lemma 2.5, $\alpha \geq \frac{\alpha'}{2}$. By Corollary 2.4, $\alpha \geq \frac{1}{4(n-1)}$, and consequently by Corollary 2.1 the mixing time for the Heavy Ace walk satisfies

$$\tau \leq 4(n-1)[3 + \log\log n!].$$

Using the method detailed in [Dia88], we can find a corresponding lower bound. For simplicity we will examine the discrete time chain K_n (the argument for H_t is analogous). Let A be the subset of permutations with at least one fixed point. That is $A = \{\sigma \in S_n \mid \sigma_i = i \text{ for some } 1 \leq i \leq n\}$. Under the uniform measure π, this is the matching problem, and arguments in [Fel68] show that

$$\pi(A) = 1 - \frac{1}{e} + O\left(\frac{1}{n!}\right).$$

Let $\{(1, X_1), (1, X_2), \ldots, (1, X_k)\}$ denote the transpositions that we considered making up to step k. That is, at step i, $1 \leq i \leq k$, we choose cards 1 and X_i, checked if either was an 'Ace of Spades', and continued accordingly. Then to bound $K_k(A)$, observe that $K_k(A) \geq K_k(B)$ where $B = \left\{\left(\bigcup_{1 \leq i \leq k} X_i\right) \neq \{2, \ldots, n\}\right\}$, i.e. B is the event that by step k we had not even chosen all of the positions. Arguments in [Fel68] show that

$$K_k(B) = 1 - e^{-ne^{-k/n}} + o(1) \qquad \text{uniformly in } k \text{ as } n \to \infty.$$

For $k = n \log n + cn$, $K_k(A) \geq 1 - e^{-e^{-c}} + o(1)$, and consequently

$$\begin{aligned} \|K_k - \pi\|_{TV} &\geq |K_k(A) - \pi(A)| \\ &\geq \frac{1}{e} - e^{-e^{-c}} + o(1). \end{aligned}$$

2.4 Concentration of Measure

In this section we present the well known connection between log Sobolev and concentration inequalities (see e.g. [Led01, BT]), and present some examples based on the inequalities derived in Section 2.2.

First we review the key definitions and results. Let (X, d, μ) denote a metric space (X, d) equipped with a probability measure μ on its Borel sets.

Definition 2.3. The concentration function on (X, d, μ) is given by

$$\alpha_{(X,d,\mu)}(r) = \sup \left\{1 - \mu(A_r) : A \subset X,\ \mu(A) \geq \frac{1}{2}\right\} \qquad r > 0$$

where $A_r = \{x \in X : d(x, A) < r\}$ is the open r-neighborhood of A with respect to the metric d.

Theorem 2.4 shows that modified log Sobolev inequalities imply deviation inequalities for Lipschitz functions. Lemma 2.6 shows that these deviation bounds in turn yield concentration inequalities; for a proof of Lemma 2.6, see [Led01].

Definition 2.4. A real-valued function F on (X, d) is said to be Lipschitz if

$$\|F\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} < \infty.$$

We say that F is 1-Lipschitz if $\|F\|_{\mathrm{Lip}} \leq 1$.

Lemma 2.6. *Let μ be a Borel probability measure on a metric space (X, d). Assume that for some non-negative, decreasing function α on $\mathbb{R}_+$ and any bounded 1-Lipschitz function F on (X, d),*

$$\mu\left(\{F \geq EF + r\}\right) \leq \alpha(r)$$

for $r > 0$. Then

$$1 - \mu(A_r) \leq \alpha(\mu(A)r)$$

for every Borel set A with $\mu(A) > 0$ and every $r > 0$. In particular,

$$\alpha_{(X,d,\mu)}(r) \leq \alpha\left(\frac{r}{2}\right).$$

For a reversible Markov chain (K, π) on state space $\mathcal{X}$, consider the graph $G = (\mathcal{X}, E)$ with symmetric edge set $E = \{(x, y) \mid \pi(x)K(x, y) > 0\}$. Then using the natural graph distance d, we can define the metric probability space $(\mathcal{X}, d, \pi)$. Theorem 2.4 follows the Herbst argument to relate the modified log Sobolev constant to a deviation inequality on this graph. For a discussion of this method and more examples, see [Led01].

Theorem 2.4. *Let α denote the modified log Sobolev constant for the reversible Markov chain (K, π) on $\mathcal{X}$. For any 1-Lipschitz function F on $(\mathcal{X}, d)$, and $r > 0$*

$$\mu\left(\{F \geq EF + r\}\right) \leq e^{-\frac{\alpha}{2}r^2}.$$

Proof. Since by definition

$$\mathcal{L}(f) \leq \frac{1}{2\alpha} E\left[\sum_y [f^2(x) - f^2(y)] \cdot [\log f^2(x) - \log f^2(y)] K(x,y)\right], \qquad (2.7)$$

letting $f^2 = e^{\lambda F - \frac{1}{2\alpha}\lambda^2}$ in (2.7), we get

$$E\left[\left(\lambda F - \frac{1}{2\alpha}\lambda^2\right) e^{\lambda F - \frac{1}{2\alpha}\lambda^2}\right] - \Lambda(\lambda)\log\Lambda(\lambda) \leq \frac{1}{2\alpha}\lambda^2\Lambda(\lambda)$$

where $\Lambda(\lambda) = Ef^2$. So,

$$\lambda\Lambda'(\lambda) \leq \Lambda(\lambda)\log\Lambda(\lambda).$$

Now let $H(\lambda) = \frac{\log\Lambda(\lambda)}{\lambda}$, with $H(0) = \frac{\Lambda'(0)}{\Lambda(0)} = EF$. Then

$$\begin{aligned} H'(\lambda) &= -\frac{\log\Lambda(\lambda)}{\lambda^2} + \frac{\Lambda'(\lambda)}{\lambda\Lambda(\lambda)} \\ &\leq 0. \end{aligned}$$

Consequently, $H(\lambda) \leq H(0)$. That is,

$$\Lambda(\lambda) \leq e^{\lambda EF}$$

and so

$$Ee^{\lambda F} \leq e^{\lambda EF + \frac{1}{2\alpha}\lambda^2}.$$

Finally, for $r > 0$

$$\begin{aligned} \mu\left(\{F \geq EF + r\}\right) &= \mu\left(\left\{e^{\lambda F} \geq e^{\lambda(EF+r)}\right\}\right) \\ &\leq e^{-\lambda(EF+r)} Ee^{\lambda F} \\ &\leq e^{\frac{1}{2\alpha}\lambda^2 - \lambda r}. \end{aligned}$$

Taking $\lambda = r\alpha$ yields the result. $\square$

Using the modified log Sobolev inequalities derived in Section 2.2, we can obtain corresponding concentration inequalities. Here we consider two examples: random-transposition and the top-random transposition shuffle.

Consider the metric probability space on the symmetric group S_n given by the random transposition metric and the uniform probability distribution π. Since by Theorem 1.1 the modified log Sobolev constant for the associated walk satisfies $\alpha \geq \frac{1}{n-1}$. Theorem 2.4 shows that for any 1-Lipschitz function on S_n, and $r > 0$,

$$\pi\left(\{F \geq EF + r\}\right) \leq e^{-\frac{1}{2(n-1)}r^2}.$$

Accordingly the concentration function satisfies

$$\alpha(r) \leq e^{-\frac{1}{8(n-1)}r^2}.$$

This random transposition graph was also studied in [BHT] via the subgaussian constant, which implies a concentration inequality. In particular, the authors show that

$$\alpha(r) \leq e^{-\frac{(2r-\sqrt{n-1})^2}{2(n-1)}}.$$

As a second example, consider the graph associated with the top-random transposition shuffle and let $\tilde{d}$ be the associated metric on S_n. Then, since $(x, y) = (1, x)(1, y)(1, x)$, $\tilde{d}(\sigma_1, \sigma_2) \leq 3d(\sigma_1, \sigma_2)$. Consequently,

$$\alpha_{\tilde{d}}(r) \leq \alpha_d(\frac{r}{3}) \leq e^{-\frac{1}{72(n-1)}r^2}.$$

However, since Corollary 2.4 shows that for this chain $\alpha \geq \frac{1}{2(n-1)}$, for any 1-Lipschitz function F on $(S_n, \tilde{d})$ and $r > 0$,

$$\pi\left(\{F \geq EF + r\}\right) \leq e^{-\frac{1}{4(n-1)}r^2}.$$

Accordingly, we have the slightly tighter bound

$$\alpha_{\tilde{d}}(r) \leq e^{-\frac{1}{16(n-1)}r^2}.$$

Chapter 3

The Spectral Profile

3.1 Background

It is well known that the spectral gap of a Markov chain can be estimated in terms of conductance, facilitating isoperimetric bounds on mixing time (see [SJ89, LS88]). Observing that small sets often have large conductance, Lovász and Kannan [LK99] strengthened this result by bounding total variation mixing time for reversible chains in terms of the "average conductance" taken over sets of various sizes. Morris and Peres [MP] introduced the idea of evolving sets to analyze reversible and non-reversible chains, and found bounds on the larger L^∞ mixing time.

To sidestep conductance, we introduce "spectral profile" and develop Faber-Krahn inequalities in the context of finite Markov chains, bounding mixing time directly in terms of the spectral profile. Faber-Krahn inequalities were introduced by Grigor'yan and developed together with Coulhon and Pittet ([Gri94, Cou96, CGP01, BCG01]) to estimate the rate of decay of the heat kernel on manifolds and infinite graphs. Their techniques build on functional analytic methods presented, for example, in [Dav90]. We adapt this approach to the setting of finite Markov chains and derive L^∞ mixing time estimates for both reversible and non-reversible walks.

These bounds let us recover the previous conductance-based results, and in general lead to sharper estimates on rates of convergence to stationarity. We also show that the spectral profile can be bounded in terms of both log-Sobolev and Nash inequalities, leading to new and elementary proofs for previous mixing time results – for example, we re-derive Theorem 3.7 of Diaconis–Saloff-Coste [DSC96a] and Theorem 42 (Chapter 8) of Aldous-Fill [AF].

In terms of applications, we first observe that for simple examples such as random walk on the complete graph and the n-cycle, the spectral profile gives the correct mixing time bounds. As more interesting examples, we show optimal bounds for walks on graphs with moderate growth, walks on the fractal-like Viscek graphs, and for the group walks on the product $Z_a \times Z_b$. In the case of graphs with moderate growth, we show that the mixing time is of the order of the square of the diameter, a result originally due to Diaconis and Saloff-Coste (see [DSC94, DSC96b]). In the case of the Viscek graphs, we show that the spectral profile provides tight upper and lower bounds on mixing time, and observe that the conductance-based bounds give much weaker upper bounds.

In Section 3.1 we introduce notation, review preliminary ideas and state our main results. Section 3.2 presents the proofs of both the continuous and discrete time versions of the spectral profile upper bound on mixing time. In Section 3.3 we recall a complementary lower bound shown in [CGP01]. Section 3.4 discusses applications, including the relationship between the spectral profile, log-Sobolev and Nash inequalities. Section 3.4.4 discusses the more elaborate example of the Viscek graphs. Section 3.4.5 discusses the spectral profile of the random walk on $Z_a \times Z_b$, which turns out to be a bit subtle. The results in this chapter are joint with Ravi Montenegro and Prasad Tetali, and can also be found in [GMT].

Our main result, Theorem 1.2, bounds the L^∞ mixing time of a chain through eigenvalues of restricted Laplace operators.

Definition 3.1. For a non-empty subset $S \subset \mathcal{X}$, define

$$\lambda(S) = \inf_{f \in c_0^+(S)} \frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)}$$

where $c_0^+(S) = \{f : \mathrm{supp}(f) \subset S,\ f \geq 0\}$.

In the reversible case,

$$\lambda_0(S) \le \lambda(S) \le \frac{1}{1-\pi(S)}\, \lambda_0(S) \tag{3.1}$$

where λ_0 is the smallest eigenvalue of the restricted Laplacian $\Delta_S : c_0(S) \to c_0(S)$ with $c_0(S) = \{f : \text{supp}(f) \subset S\}$ and

$$\Delta_S f(x) = \begin{cases} \Delta f(x) & x \in S \\ 0 & x \notin S \end{cases}$$

The kernel of $\Delta_S = I - K_S$ is given explicitly by

$$K_S(x,y) = \begin{cases} K(x,y) & x, y \in S \\ 0 & \text{otherwise} \end{cases}$$

By the Courant-Fischer minmax characterization of eigenvalues, (3.1) is equivalent to the statement:

$$\inf_{f \in c_0(S)} \frac{\mathcal{E}_{K_s}(f,f)}{\|f\|_2^2} \le \inf_{f \in c_0^+(S)} \frac{\mathcal{E}_K(f,f)}{\text{Var}(f)} \le \frac{1}{1-\pi(S)} \inf_{f \in c_0(S)} \frac{\mathcal{E}_{K_S}(f,f)}{\|f\|_2^2}$$

The lower bound is due to the identity $\mathcal{E}_K(f,f) = \mathcal{E}_{K_S}(f,f)$ when $f \in c_0(S)$, which follows from $\Delta f(x) = \Delta_S f(x)$ when $f \in c_0(S)$ and $x \in S$. The upper bound also requires the inequality $(x-y)^2 \ge (|x|-|y|)^2$ to show that $\mathcal{E}(f,f) \ge \mathcal{E}(|f|,|f|)$, while Cauchy-Schwartz gives $\|f\|_1 \le \|f\|_2 \sqrt{\pi(S)}$ which implies $\text{Var}(f) \ge (1-\pi(S))\|f\|_2^2$. In general, when $\pi(S) \le 1/2$ then $\lambda(S)$ is within a factor two of the smallest eigenvalue of the symmetric operator $(\Delta_S + \Delta_S^*)/2$.

We are interested in how $\lambda(S)$ decays as the size of S increases.

Definition 3.2. Define the spectral profile $\Lambda : [\pi_*, \infty) \to \mathbb{R}$ by

$$\Lambda(r) = \inf_{\pi_* \le \pi(S) \le r} \lambda(S).$$

Observe that $\Lambda(r)$ is non-increasing, and $\Lambda(r) \geq \lambda_1$. Furthermore, by construction the walk (K, π) satisfies the *Faber-Krahn inequality*

$$\lambda(S) \geq \Lambda(\pi(S)) \quad \forall S \subset \mathcal{X}\,.$$

Theorem 1.2 is our main result:

Theorem 1.2. For $\epsilon > 0$, the L^∞ mixing time $\tau_\infty(\epsilon)$ for a chain $H_t(x, y)$ satisfies

$$\tau_\infty(\epsilon) \leq \int_{4\pi_*}^{4/\epsilon} \frac{2dv}{v\Lambda(v)}\,.$$

In Section 3.2.3, we prove an analogous result for discrete-time walks. Since $\Lambda(r) \geq \lambda_1$, Theorem 1.2 shows that

$$\tau_\infty(1/e) \leq \int_{4\pi_*}^{4e} \frac{2\,dv}{v\Lambda(v)} \leq \frac{2}{\lambda_1}\left(1 + \log\frac{1}{\pi_*}\right).$$

But since we can expect $\Lambda(r) \gg \lambda_1$ for small r, Theorem 1.2 offers an improvement over the standard spectral gap mixing time bound (1.2).

By a discrete version of the Cheeger inequality of differential geometry,

$$\Phi_*^2(r)/2 \leq \Lambda(r) \leq 2\Phi_*(r)$$

where $\Phi_*(r)$ is the (truncated) conductance profile (see Section 3.2.2). Consequently, by Theorem 1.2:

Corollary 1.2. For $\epsilon > 0$, the L^∞ mixing time $\tau_\infty(\epsilon)$ for a chain $H_t(x, y)$ satisfies

$$\tau_\infty(\epsilon) \leq \int_{4\pi_*}^{4/\epsilon} \frac{4dv}{v\Phi_*^2(v)}.$$

Theorem 13 of [MP] is a factor of two weaker than this. Although Theorem 1.2 implies mixing time estimates in terms of conductance, it is reasonable to expect that for many models $\Lambda(r) \gg \Phi_*^2(r)$. In these cases, compared to Corollary 1.2, presently the best known conductance bound, our spectral approach leads to sharper mixing time results. We provide below examples of such cases (see Sections 3.4.1 and 3.4.4).

3.2 Upper Bounds on Mixing Time

3.2.1 Spectral Profile Bounds

In this section, we prove one of the main results, Theorem 1.2. The proof uses the techniques of [Gri94] for estimating heat kernel decay on non-compact manifolds. The first Dirichlet eigenvalue $\lambda_0(S)$ for small sets S captures the convergence behavior at the start of the walk, when the fact that the state space is finite has minimal influence. The spectral gap λ_1 governs the long-term convergence. The spectral profile $\Lambda(r)$ takes into account these two effects, since $\lambda(S) \approx \lambda_0(S)$ for $\pi(S) \leq 1/2$, and by Lemma 3.2 $\Lambda(r) \approx \lambda_1$ for $r \geq 1/2$.

To bound mixing times, we first lower bound $\mathcal{E}(f, f)$ in terms of the spectral profile Λ, and as such Lemma 3.1 is the crucial step in the proof of Theorem 1.2.

We regularly use the notation that, given a function f, $f_+ = f \vee 0$ denotes its positive part, and $f_- = -(f \wedge 0)$ its negative part.

Lemma 3.1. *For every non-constant function $u : \mathcal{X} \mapsto \mathbb{R}_+$,*

$$\frac{\mathcal{E}(u,u)}{\operatorname{Var} u} \geq \frac{1}{2}\,\Lambda\Big(\frac{4(\mathbb{E}u)^2}{\operatorname{Var} u}\Big)\,.$$

Proof. For c constant, $\mathcal{E}(u,u) = \mathcal{E}(u-c, u-c)$. Also, $\forall a, b \in \mathbb{R} : (a-b)^2 \geq (a_+ - b_+)^2$ so $\mathcal{E}(f,f) \geq \mathcal{E}(f_+, f_+)$. It follows that when $0 \leq c < \max u$ then

$$\begin{aligned}
\mathcal{E}(u,u) &\geq \mathcal{E}((u-c)_+, (u-c)_+) \\
&\geq \operatorname{Var}((u-c)_+) \inf_{f \in c_0^+(u>c)} \frac{\mathcal{E}(f,f)}{\operatorname{Var}(f)} \\
&\geq \operatorname{Var}((u-c)_+)\,\Lambda(\pi(u>c))\,.
\end{aligned}$$

Now, $\forall a, b \geq 0 :\ (a-b)_+^2 \geq a^2 - 2b\,a$ and $(a-b)_+ \leq a$ so

$$\operatorname{Var}((u-c)_+) = \mathbb{E}(u-c)_+^2 - (\mathbb{E}(u-c)_+)^2 \geq \mathbb{E}u^2 - 2c\,\mathbb{E}u - (\mathbb{E}u)^2.$$

Let $c = \mathrm{Var}(u)/4\mathbb{E}u$ and observe that

$$c \le \frac{\mathbb{E}u^2}{4\mathbb{E}u} \le \frac{\max u}{4}.$$

Applying Markov's inequality $\pi(u > c) < (\mathbb{E}u)/c$,

$$\mathcal{E}(u,u) \ge (\mathrm{Var}(u) - 2c\,\mathbb{E}u)\,\Lambda(\mathbb{E}u/c) = \frac{1}{2}\,\mathrm{Var}(u)\,\Lambda\left(\frac{4(\mathbb{E}u)^2}{\mathrm{Var}\,u}\right).$$

□

Now we bound the L^2 distance of a chain from equilibrium in terms of the function $V(t) : [0,\infty) \to \mathbb{R}$ given by

$$t = \int_{4\pi_*}^{V(t)} \frac{dv}{v\Lambda(v)}.$$

Since the integral diverges, $V(t)$ is well-defined for $t \ge 0$.

The L^2 bound of Theorem 3.1 implies the L^∞ bound that is our main result. To prove the L^2 bound, we simply apply Lemma 3.1 to the heat kernel $h(x,y,t)$.

Theorem 3.1. *For the chain (K,π), we have*

$$\sup_{x\in\mathcal{X}} d^2_{\pi,2}(H^x_t,\pi) \le \frac{4}{V(t)}.$$

Proof. Given $x \in \mathcal{X}$ a value where the supremum occurs, define $u_{x,t}(y) = h(x,y,t)$ and $I_x(t) = \mathrm{Var}(u_{x,t})$. If $u_{x,t} = 1$ then the theorem follows trivially. Otherwise, $u_{x,t}$ is non-constant and since $\mathbb{E}u_{x,t} = 1$, by Proposition 1.4 and Lemma 3.1

$$I'_x(t) = -2\mathcal{E}(u_{x,t},u_{x,t}) \le -I_x\Lambda(4/I_x). \tag{3.2}$$

Integrating over $[0,t]$ we have

$$\int_{I_x(0)}^{I_x(t)} \frac{dI_x}{I_x\Lambda(4/I_x)} \le -t.$$

With the change of variable $v = 4/I_x$,

$$t \le \int_{4/I_x(0)}^{4/I_x(t)} \frac{dv}{v\Lambda(v)}.$$

Since $I_x(0) = 1/\pi(y) - 1 < 1/\pi_*$

$$V(t) \le \frac{4}{I_x(t)} = \frac{4}{\|h(x,\cdot,t) - 1\|_2^2}$$

and the result follows. □

Now we show how to transfer the L^2 bounds of Theorem 3.1 to the L^∞ bounds of our main result.

Proof of Theorem 1.2. Observe that

$$\begin{aligned} \left|\frac{H_t(x,y) - \pi(y)}{\pi(y)}\right| &= \left|\frac{\sum_z \left(H_{t/2}(x,z) - \pi(z)\right)\left(H_{t/2}(z,y) - \pi(y)\right)}{\pi(y)}\right| \\ &= \left|\sum_z \pi(z)\left(\frac{H_{t/2}(x,z)}{\pi(z)} - 1\right)\left(\frac{H^*_{t/2}(y,z)}{\pi(z)} - 1\right)\right| \\ &\le d_{\pi,2}(H_{t/2}(x,\cdot),\pi)\, d_{\pi,2}(H^*_{t/2}(y,\cdot),\pi) \end{aligned} \tag{3.3}$$

where the inequality follows from Cauchy-Schwartz. Since we can apply Theorem 3.1 to either H_t or H_t^*, we have

$$\sup_{x,y\in\mathcal{X}} |h(x,y,t) - 1| \le \frac{4}{V(t/2)}.$$

So $|h(x,y,t) - 1| \le \epsilon$ for $V(t/2) \ge 4/\epsilon$, that is, for t such that

$$t/2 \ge \int_{4\pi_*}^{4/\epsilon} \frac{dv}{v\Lambda(v)}$$

proving the result. □

The next result shows that any improvement in using the spectral profile $\Lambda(r)$ instead of the spectral gap λ_1 comes from looking at small sets since for $r = 1/2$, already $\Lambda(r) \approx \lambda_1$.

Lemma 3.2. *The spectral gap* λ_1 *and the spectral profile* $\Lambda(r)$ *satisfy*

$$\lambda_1 \leq \Lambda(1/2) \leq 2\lambda_1.$$

Proof. The lower bound follows immediately from the definition of the spectral gap. For the upper bound, let m be a median of f. Then using Lemma 3.3,

$$\begin{aligned} \mathcal{E}(f,f) &= \mathcal{E}(f-m, f-m) \\ &\geq \mathcal{E}((f-m)_+, (f-m)_+) + \mathcal{E}((f-m)_-, (f-m)_-). \end{aligned}$$

Since $\pi(\{f > m\}) = \pi(\{f < m\}) \leq 1/2$, we have

$$\mathcal{E}((f-m)_+, (f-m)_+) \geq \|(f-m)_+\|_2^2 \, \lambda_0(\{f > m\})$$

and

$$\mathcal{E}((f-m)_-, (f-m)_-) \geq \|(f-m)_-\|_2^2 \, \lambda_0(\{f < m\})$$

Consequently,

$$\begin{aligned} \mathcal{E}(f,f) &\geq \|f-m\|_2^2 \inf_{\pi(S) \leq 1/2} \lambda_0(S) \\ &\geq \mathrm{Var}(f) \, \frac{\Lambda(1/2)}{2} \, . \end{aligned}$$

The upper bound follows by minimizing over f. □

The proof required the following lemma.

Lemma 3.3. *Given a function* $f : \mathcal{X} \mapsto \mathbb{R}$ *then*

$$\mathcal{E}(f,f) \geq \mathcal{E}(f_+, f_+) + \mathcal{E}(f_-, f_-) \geq \mathcal{E}(|f|, |f|).$$

Proof. Given $g, h : \mathcal{X} \mapsto \mathbb{R}$ with $g, h \geq 0$ and $(\operatorname{supp} g) \cap (\operatorname{supp} h) = \emptyset$ then

$$\mathcal{E}(g,h) = \sum_x g(x)h(x)\pi(x) - \sum_{x,y} g(y)h(x)K(x,y)\pi(x) \leq 0$$

because the first sum is zero and every term in the second is non-negative. In particular, $f_+, f_- \geq 0$ with $(\operatorname{supp} f_+) \cap (\operatorname{supp} f_-) = \emptyset$, and so by linearity

$$\begin{aligned}
\mathcal{E}(f,f) &= \mathcal{E}(f_+ - f_-, f_+ - f_-) \\
&= \mathcal{E}(f_+, f_+) + \mathcal{E}(f_-, f_-) - \mathcal{E}(f_+, f_-) - \mathcal{E}(f_-, f_+) \\
&\geq \mathcal{E}(f_+, f_+) + \mathcal{E}(f_-, f_-) \\
&\geq \mathcal{E}(f_+, f_+) + \mathcal{E}(f_-, f_-) + \mathcal{E}(f_+, f_-) + \mathcal{E}(f_-, f_+) \\
&= \mathcal{E}(|f|, |f|).
\end{aligned}$$

$\square$

3.2.2 Conductance Bounds

In this section, we show how to use Theorem 1.2 to recover previous bounds on mixing time in terms of the conductance profile.

Definition 3.3. For non-empty $A, B \subset \mathcal{X}$, the flow is given by

$$Q(A,B) = \sum_{x \in A,\, y \in B} Q(x,y)$$

where $Q(x,y) = \pi(x)K(x,y)$ can be viewed as a probability measure on $\mathcal{X} \times \mathcal{X}$. The boundary of a subset is defined by

$$\partial S = \{x \in S : \exists y \notin S,\ K(x,y) > 0\}$$

and $|\partial S| = Q(S, S^c)$.

Observe that

$$\pi(S) = Q(S, \mathcal{X}) = Q(S,S) + Q(S, S^c)$$

and also

$$\pi(S) = Q(\mathcal{X}, S) = Q(S,S) + Q(S^c, S).$$

It follows that $Q(S, S^c) = Q(S^c, S)$.

Like the spectral profile $\Lambda(r)$, the conductance profile $\Phi(r)$ measures how conductance changes with the size of the set S.

Definition 3.4. Define the conductance profile $\Phi : [\pi_*, 1) \to \mathbb{R}$ by

$$\Phi(r) = \inf_{\pi_* \leq \pi(S) \leq r} \frac{|\partial S|}{\pi(S)}$$

and the truncated conductance profile $\Phi_* : [\pi_*, 1) \to \mathbb{R}$ by

$$\Phi_*(r) = \begin{cases} \Phi(r) & r < 1/2 \\ \Phi(1/2) & r \geq 1/2 \end{cases}$$

The value $\Phi(1/2)$ is often referred to as the conductance, or the isoperimetric constant, of the chain.

The next lemma is a discrete version of the "Cheeger inequality" of differential geometry, and will let us apply Theorem 1.2 to recover the conductance profile bound of Corollary 1.2. The proof of the lemma is similar to the proof given in [SC96] of the fact that

$$\frac{\Phi^2(1/2)}{8} \leq \lambda_1 \leq 2\Phi(1/2).$$

Lemma 3.4. *For $r \in [\pi_*, 1)$, the spectral profile Λ and the conductance profile Φ satisfy*

$$\frac{\Phi^2(r)}{2} \leq \Lambda(r) \leq \frac{\Phi(r)}{1-r}.$$

Proof. It suffices to show that $\frac{1}{2}\Phi^2(\pi(A)) \leq \lambda_0(A) \leq \frac{|\partial A|}{\pi(A)}$ for every $A \subset \mathcal{X}$. The bound then follows from (3.1) by minimizing over sets with $\pi(A) \leq r$.

For the upper bound,

$$\lambda_0(A) \leq \frac{\mathcal{E}(1_A, 1_A)}{\|1_A\|_2^2} = \frac{|\partial A|}{\pi(A)}.$$

To show the lower bound, for a non-negative function f, define the level sets $F_t = \{x \in \mathcal{X} : f(x) \geq t\}$ and the indicator functions $f_t = 1_{F_t}$. Then

$$\begin{aligned} \pi(f) &= \sum_{x \in \mathcal{X}} \left(\int_0^\infty f_t(x) dt \right) \pi(x) \qquad (3.4) \\ &= \int_0^\infty \pi(F_t)\, dt. \end{aligned}$$

Furthermore,

$$\begin{aligned} \sum_{x,y} |f(x) - f(y)|\, Q(x,y) &= \frac{1}{2} \sum_{x,y} |f(x) - f(y)|\, [Q(x,y) + Q(y,x)] \\ &= \sum_{f(x) > f(y)} [f(x) - f(y)]\, [Q(x,y) + Q(y,x)] \\ &= \sum_{f(x) > f(y)} \int_0^\infty 1_{\{f(y) < t \leq f(x)\}}\, [Q(x,y) + Q(y,x)]\, dt \\ &= \int_0^\infty |\partial F_t|\, dt + \int_0^\infty |\partial F_t^c|\, dt \\ &= 2 \int_0^\infty |\partial F_t|\, dt. \qquad (3.5) \end{aligned}$$

Observe that (3.5) is a discrete analog of the co-area formula. For non-negative $f \in c_0(A)$, $F_t \subset A$ for $t > 0$, and so

$$\begin{aligned} \sum_{x,y} |f(x) - f(y)| Q(x,y) &= 2 \int_0^\infty |\partial F_t|\, dt \qquad \text{by (3.5)} \\ &\geq 2\Phi(\pi(A)) \int_0^\infty \pi(F_t)\, dt \\ &= 2\Phi(\pi(A))\pi(f) \qquad \text{by (3.4)}. \end{aligned}$$

Consequently, for any non-negative $f \in c_0(A)$, by the above

$$\begin{aligned}
2\Phi(\pi(A))\pi(f^2) &\leq \sum_{x,y} |f^2(x) - f^2(y)|\, Q(x,y) \\
&= \sum_{x,y} |f(x) - f(y)| \cdot (f(x) + f(y))\, Q(x,y) \\
&\leq \left(\sum_{x,y} (f(x) - f(y))^2\, Q(x,y) \right)^{1/2} \\
&\qquad \times \left(\sum_{x,y} (f(x) + f(y))^2\, Q(x,y) \right)^{1/2} \\
&\leq (2\mathcal{E}(f,f))^{1/2}\, (4\pi(f^2))^{1/2}.
\end{aligned}$$

Then

$$\lambda_0(A) = \inf_{f \in c_0^+(A)} \frac{\mathcal{E}(f,f)}{\pi(f^2)} \geq \frac{\Phi^2(\pi(A))}{2}\,.$$

The infimum for $\lambda_0(A)$ occurred at $f \in c_0^+(A)$ because for general $f \in c_0(A)$, $\mathcal{E}(f,f) \geq \mathcal{E}(|f|,|f|)$ and $\pi(f^2) = \pi(|f|^2)$. $\square$

From the proofs of Lemma 3.4 and Lemma 3.2, we have that

$$\frac{\Phi^2(1/2)}{2} \leq \inf_{\pi(A) \leq 1/2} \lambda_0(A) \leq \lambda_1.$$

Consequently, $\Phi_*^2(r) \leq 2\Lambda(r)$, proving the conductance profile bound of Corollary 1.2.

3.2.3 Discrete-Time Walks

In this section we consider discrete-time chains, deriving spectral profile bounds on mixing time similar to those for continuous-time walks. For $u_{x,t}(y) = h(x,y,t)$, the rate of decay of the heat operator in the continuous setting is given by

$$\frac{d}{dt}\mathrm{Var}(u_{x,t}) = -2\mathcal{E}(u_{x,t}, u_{x,t}). \tag{3.6}$$

In the discrete-time setting, set

$$u_{x,n}(y) = k(x,y,n) = \frac{K_n(x,y)}{\pi(y)}.$$

Then, since $K^* u_{x,n} = u_{x,n+1}$ and $\mathbb{E}(u_{x,n}) = 1$,

$$\begin{aligned} \mathrm{Var}(u_{x,n+1}) - \mathrm{Var}(u_{x,n}) &= \langle u_{x,n+1}, u_{x,n+1}\rangle - \langle u_{x,n}, u_{x,n}\rangle \\ &= -\langle (I - KK^*)u_{x,n},\ u_{x,n}\rangle \\ &= -\mathcal{E}_{KK^*}(u_{x,n}, u_{x,n}) \end{aligned} \tag{3.7}$$

and so it is natural to consider the multiplicative symmeterizations KK^* and K^*K. In order to relate mixing time directly to the kernel K of the original walk, we use the assumption that for $\alpha > 0$

$$K(x,x) \geq \alpha \quad \forall x \in \mathcal{X}.$$

Define Λ_{KK^*} and V_{KK^*} to be the analogs of Λ and V where $\mathcal{E}_K(f,f)$ is replaced by $\mathcal{E}_{KK^*}(f,f)$. If KK^* is reducible, then $\lambda_1^{KK^*} = 0$, and so we restrict ourselves to the irreducible case. We define Λ_{K^*K} and V_{K^*K} similarly and also assume irreducibility. The following result is a discrete-time version of Theorem 3.1, and its proof is analogous.

Theorem 3.2. *For a discrete-time chain (K,π) with K^*K and KK^* irreducible*

$$\sup_{x\in\mathcal{X}} d^2_{\pi,2}(K_n(x,\cdot),\pi) \leq \frac{4}{V_{KK^*}(n/2)} \quad \textit{and} \quad \sup_{x\in\mathcal{X}} d^2_{\pi,2}(K^*_n(x,\cdot),\pi) \leq \frac{4}{V_{K^*K}(n/2)}.$$

Proof. The second statement follows from the first by replacing K by K^*. For fixed $x \in \mathcal{X}$, define $u_{x,n}(y) = k(x,y,n)$ and $I_x(n) = \mathrm{Var}(u_{x,n})$. By (3.7) and Lemma 3.1

$$\begin{aligned} I_x(n+1) - I_x(n) &= -\mathcal{E}_{KK^*}(u_{x,n}, u_{x,n}) \\ &\leq -\frac{1}{2} I_x(n)\Lambda_{KK^*}(4/I_x(n)). \end{aligned}$$

Since both $I_x(n)$ and $\Lambda_{KK^*}(r)$ are non-increasing, the piecewise linear extension of $I_x(n)$ to $\mathbb{R}_+$ satisfies

$$I'_x(t) \leq -\frac{1}{2} I_x(t) \Lambda_{KK^*}(4/I_x(t)).$$

At integer t, we can take either the derivative from the right or the left. Solving this differential equation as in Theorem 3.1, we have

$$V_{KK^*}(t/2) \leq \frac{4}{I_x(t)}$$

and the result follows. □

Corollary 3.1. *Assume that $K(x,x) \geq \alpha > 0$ for all $x \in \mathcal{X}$. Then for $\epsilon > 0$, the L^∞ mixing time for the discrete-time chain K_n satisfies*

$$\tau_\infty(\epsilon) \leq 2 \left\lceil \int_{4\pi_*}^{4/\epsilon} \frac{dv}{\alpha v \Lambda(v)} \right\rceil .$$

Proof. Since $K^*(x,x) = K(x,x) \geq \alpha$, observe that

$$\begin{aligned} KK^*(x,y)\pi(x) &\geq K^*(x,x)K(x,y)\pi(x) + K^*(x,y)K(y,y)\pi(x) \\ &\geq \alpha K(x,y)\pi(x) + \alpha K(y,x)\pi(y) \end{aligned}$$

and so,

$$\mathcal{E}_{KK^*}(f,f) \geq 2\alpha \mathcal{E}_K(f,f).$$

Consequently, $\Lambda_{KK^*} \geq 2\alpha\Lambda$, from which it follows that

$$\begin{aligned} \alpha t &= \int_{4\pi_*}^{V(\alpha t)} \frac{dv}{v\Lambda(v)} \\ &\geq 2\alpha \int_{4\pi_*}^{V(\alpha t)} \frac{dv}{v\Lambda_{KK^*}(v)}. \end{aligned}$$

Accordingly, $V_{KK^*}(t/2) \geq V(\alpha t)$, and similarly $V_{K^*K}(t/2) \geq V(\alpha t)$. As in Theorem 1.2,

$$\begin{aligned} |k(x,y,2n) - 1| &\leq d_{\pi,2}(K_n(x,\cdot),\pi)\, d_{\pi,2}(K_n^*(y,\cdot),\pi) \\ &\leq \frac{4}{V(\alpha n)}. \end{aligned}$$

And so, $|k(x,y,2n)-1| \le \epsilon$ for

$$n \ge \int_{4\pi_*}^{4/\epsilon} \frac{dv}{\alpha v \Lambda(v)}.$$

$\square$

Improvement for discrete-time using rescaling. Given a Markov kernel K let $\Lambda_K(r)$ and $\Phi_K(r)$ denote the spectral and conductance profiles, respectively. Then

$$\begin{aligned}\Lambda_K(r) = (1-\alpha)\Lambda_{\frac{K-\alpha I}{1-\alpha}}(r) &\ge \frac{1-\alpha}{2}\,\Phi_{\frac{K-\alpha I}{1-\alpha}}(r)^2 \\ = \frac{1-\alpha}{2}\left(\frac{\Phi_K(r)}{1-\alpha}\right)^2 &= \frac{\Phi_K(r)^2}{2(1-\alpha)}.\end{aligned}$$

The appropriate discrete time version of Corollary 1.2 is then

Corollary 3.2. *For $\epsilon > 0$, the L^∞ mixing time $\tau_\infty(\epsilon)$ for the chain K_n satisfies*

$$\tau_\infty(\epsilon) \le 2\left\lceil \int_{4\pi_*}^{4/\epsilon} \frac{2\,dv}{\frac{\alpha}{1-\alpha}\,v\Phi_*^2(v)} \right\rceil.$$

In contrast, the bound of Morris and Peres [MP] is

$$\tau_\infty(\epsilon) \le 2\left\lceil \int_{4\pi_*}^{4/\epsilon} \frac{2\,dv}{\min\left\{\frac{\alpha^2}{(1-\alpha)^2}, 1\right\} v\Phi_*^2(v)} \right\rceil$$

which is similar, but slightly weaker when $\alpha \neq 1/2$.

3.3 Lower Bounds on Mixing Time

In this section, we recall a result of [CGP01] to show that for reversible chains the spectral profile describes well the decay behavior of the heat kernel $h_t(x,y) = H_t(x,y)/\pi(y)$. These results are based on the idea of *anti-Faber-Krahn inequalities.*

For reversible chains, (3.3) implies

$$\sup_{x,y} \frac{H_t(x,y)}{\pi(y)} - 1 \le \sup_x \sum_z \pi(z) \left(\frac{H_{t/2}(x,z)}{\pi(z)} - 1 \right)^2 = \sup_x \frac{H_t(x,x)}{\pi(x)} - 1$$

and so

$$\sup_{x,y \in \mathcal{X}} h_t(x,y) = \sup_{x \in \mathcal{X}} h_t(x,x)\,.$$

Lemma 3.5 gives a simple lower bound on the heat kernel.

Lemma 3.5 ([CGP01]). *For a reversible chain (K, π) and non-empty $S \subset \mathcal{X}$,*

$$\sup_{x \in \mathcal{X}} h_t(x,x) \ge \frac{\exp(-t\lambda_0(S))}{2\pi(S)}.$$

Proof. Let $\lambda_0(S) \le \lambda_1(S) \le \cdots \le \lambda_{|S|-1}(S)$ be the eigenvalues of $I - K_S$. Then K_S has eigenvalues $\{1 - \lambda_i(S)\}$. Since $\mathrm{tr}(K_S^k)$ can be written as either the sum of eigenvalues, or the sum of diagonal entries, we have

$$\begin{aligned} \sum_{i=0}^{|S|-1} (1 - \lambda_i(S))^k &= \sum_{x \in S} K_S^k(x,x) \\ &\le \sum_{x \in S} K_k(x,x). \end{aligned}$$

For k even, all the terms in the first sum are non-negative, and consequently

$$(1 - \lambda_0(S))^k \le \sum_{x \in S} K_k(x,x).$$

Finally, to bound the continuous-time kernel, note that

$$\begin{aligned} \pi(S) \sup_{x \in \mathcal{X}} h_t(x,x) &\ge \sum_{x \in S} h_t(x,x)\pi(x) \\ &\ge \sum_{x \in S} e^{-t} \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} K_{2k}(x,x) \\ &\ge e^{-t} \sum_{k=0}^{\infty} \frac{t^{2k}(1 - \lambda_0(S))^{2k}}{(2k)!} \\ &= e^{-t} \frac{\exp[t(1 - \lambda_0(S))] + \exp[-t(1 - \lambda_0(S))]}{2} \end{aligned}$$

from which the result follows. $\square$

Theorem 3.3 is a partial converse of the upper bound given in Theorem 3.1 under the restriction of δ-regularity.

Definition 3.5. A positive, increasing function $f \in C^1(0,T)$ is δ-regular if for all $0 < t < s \leq 2t < T$

$$\frac{f'(s)}{f(s)} \geq \delta \frac{f'(t)}{f(t)}.$$

Definition 3.6. The walk (K,π) satisfies the anti-Faber-Krahn inequality with function $L : [\pi_*, \infty) \to \mathbb{R}$ if for all $r \in [\pi_*, \infty)$,

$$\inf_{\pi_* \leq \pi(S) \leq r} \lambda_0(S) \leq L(r).$$

Remark 3.1. Observe that (K,π) satisfies the anti-Faber-Krahn inequality with $L(r) = \Lambda(r)$, in light of (3.1).

Theorem 3.3 ([CGP01]). *Let (K,π) be a reversible Markov chain that satisfies the anti-Faber-Krahn inequality with $L : (\pi_*, \infty) \to \mathbb{R}$, and that $\gamma(t)$, defined implicitly by*

$$t = \int_{\pi_*}^{\gamma(t)} \frac{dv}{vL(v)},$$

is δ-regular on $(0,T)$. Then for $t \in (0, \delta T/2)$

$$\sup_{x \in \mathcal{X}} h_t(x,x) \geq \frac{1}{2\gamma(2t/\delta)}.$$

Proof. Fix $t \in (0, \delta T/2)$ and set $r = \gamma(t/\delta)$. By the anti-Faber-Krahn inequality, there exists $S \subset \mathcal{X}$ with $\pi(S) \leq r$ and $\lambda_0(S) \leq L(r)$. Consequently, by Lemma 3.5,

$$\sup_{x \in \mathcal{X}} h_t(x,x) \geq \frac{\exp(-t\lambda_0(S))}{2\pi(S)} \geq \frac{\exp(-tL(r))}{2r}.$$

So, $\sup_x h_t(x,x) \geq \exp(-C_t)$ for $C_t = \log 2r + tL(r)$. Since $L(\gamma(s)) = (\log \gamma)'(s)$

$$C_t = \log 2\gamma(t/\delta) + t(\log \gamma)'(t/\delta).$$

By the mean value theorem, there exists $\theta \in (t/\delta, 2t/\delta)$ such that

$$(\log \gamma)'(\theta) = \frac{\log \gamma(2t/\delta) - \log \gamma(t/\delta)}{t/\delta}.$$

By δ-regularity

$$(\log \gamma)'(\theta) \geq \delta (\log \gamma)'(t/\delta)$$

and so $C_t \leq \log[2\gamma(2t/\delta)]$, showing the result. □

3.4 Applications

The following lemma, while hardly surprising, is often effective in reducing computation in specific examples. In particular, it is used in computing the spectral profile of the random walk on the n-cycle in the present section.

Lemma 3.6. *Let $S = S_1 \cup \cdots \cup S_k$ be a decomposition of S into connected components. Then*

$$\lambda(S) = \min_{S_i} \{\lambda(S_i)\}.$$

Proof. Clearly $\lambda(S) \leq \min_{S_i}\{\lambda(S_i)\}$, and we need only show the reverse inequality. For a function $f \geq 0$, define $f_{S_i} = 1_{S_i} f$. Then

$$\mathrm{Var}(f) = \mathrm{Var}\left(\sum_{S_i} f_{S_i}\right) = \sum_{S_i} \mathbb{E} f_{S_i}^2 - \left(\sum_{S_i} \mathbb{E} f_{S_i}\right)^2 \leq \sum_{S_i} \mathrm{Var}(f_{S_i}).$$

Consequently,

$$\begin{aligned}
\lambda(S) &= \inf_{f \in c_0^+(S)} \frac{\mathcal{E}(f,f)}{\mathrm{Var}(f)} \\
&= \inf_{f \in c_0^+(S)} \frac{\sum_{S_i} \mathcal{E}(f_{S_i}, f_{S_i})}{\mathrm{Var}(f)} \\
&\geq \inf_{f \in c_0^+(S)} \frac{\sum_{S_i} \lambda(S_i) \mathrm{Var}(f_{S_i})}{\mathrm{Var}(f)}
\end{aligned}$$

and the result follows. □

3.4.1 First Examples

The Complete Graph. Consider the continuous-time walk on the complete graph in the n-point space $\Omega = \{x_1, \ldots, x_n\}$ with kernel $K(x_i, x_j) = 1/n \; \forall i, j$. To find the eigenvalues of the restricted operator $K_S : c_0(S) \mapsto c_0(S)$, we consider functions $f : \{x_1, \ldots, x_{|S|}\} \mapsto \mathbb{R}$. Since

$$K_S f(x_j) = \frac{1}{n} \sum_{i=1}^{|S|} f(x_i) = \bar{f} \qquad 1 \leq j \leq |S|$$

f is an eigenfunction of K_S with corresponding eigenvalue λ if and only if $\lambda f(x_j) = \bar{f}$ for $1 \leq j \leq |S|$. If $\lambda \neq 0$, then this implies that f is constant with eigenvalue $\lambda = |S|/n$. So, the smallest eigenvalue of $I - K_S$ satisfies $\lambda_0(S) = 1 - |S|/n$, and the second smallest eigenvalue of $I - K$ satisfies $\lambda_1 = 1$. Since

$$\lambda_1 \leq \lambda(S) \leq \frac{\lambda_0(S)}{1 - \pi(S)}$$

$\lambda(S) = 1$ and accordingly $\Lambda(r) \equiv 1$.

Theorem 1.2 then shows that for the complete graph $\tau_\infty(\epsilon) \leq 2 \log(n/\epsilon)$. Since the distribution of the chain at any time $t \geq 0$ is given explicitly by

$$H_t(x_i, x_j) = e^{-t} \delta_{x_i}(x_j) + \frac{(1 - e^{-t})}{n}$$

we see that $\tau_\infty(\epsilon) = \log[(n-1)/\epsilon]$, and so our estimate is off by a factor of 2.

The n-Cycle. Now consider simple random walk on the n-cycle $\{x_0, \ldots x_{n-1}\}$ given by kernel $K(x_i, x_j) = 1/2$ if $j = i \pm 1 \pmod{n}$ and zero otherwise. By Lemma 3.6, to find $\lambda_0(S)$ we need only consider connected subsets $S \subset \Omega$. For S such that $\pi(S) < 1$, $I - K_S$ corresponds to the tridiagonal Toeplitz matrix with 1's along the diagonal and 1/2's along the upper and lower off-diagonals (and 0's everywhere else). In this case, the least eigenvalue is given explicitly by

$$\lambda_0(S) = 1 - \cos\left(\frac{\pi}{|S| + 1}\right).$$

Since the spectral gap satisfies $\lambda_1 = 1 - \cos(2\pi/n)$, we have $\Lambda(r) \approx 1/(rn)^2$ for $1/n \le r \le 1$. Theorem 1.2 then shows the correct $O(n^2)$ mixing time bound.

3.4.2 Log-Sobolev and Nash Inequalities

Logarithmic Sobolev and Nash inequalities are among the strongest tools available to study L^2 convergence rates of finite Markov chains. Log-Sobolev inequalities were introduced by Gross [Gro75, Gro93] to study Markov semigroups in infinite dimensional settings, and developed in the discrete setting by Diaconis and Saloff-Coste [DSC96a]. Nash inequalities were originally formulated to study the decay of the heat kernel in certain parabolic equations (see [Nas58]). Building on ideas in [CKS87, CSC90b, CSC90a], Diaconis and Saloff-Coste [DSC96b] show how to apply Nash's argument to finite Markov chains. In this section we show that both log-Sobolev and Nash inequalities yield bounds on the spectral profile $\Lambda(r)$, leading to new proofs of previous mixing time estimates in terms of these inequalities.

Definition 3.7. The log-Sobolev constant ρ is given by

$$\rho = \inf_{\mathrm{Ent}_\pi f^2 \neq 0} \frac{\mathcal{E}(f,f)}{\mathrm{Ent}_\pi f^2}$$

where the entropy $\mathrm{Ent}_\pi(f^2) = \sum_{x\in\mathcal{X}} f^2(x) \log\left(f^2(x)/\|f\|_2^2\right) \pi(x)$.

Lemma 3.7. *The spectral profile $\Lambda(r)$ and log-Sobolev constant ρ satisfy*

$$\Lambda(r) \geq \rho \, \frac{\log(1/r)}{1-r}\,.$$

Proof. By definition

$$\Lambda(r) = \inf_{\pi(S)\le r} \inf_{f \in c_0^+(S)} \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)} \geq \rho \inf_{\pi(S)\le r} \inf_{f \in c_0^+(S)} \frac{\mathrm{Ent}_\pi(f^2)}{\mathrm{Var}_\pi(f)}$$

The lemma will follow if for every set $S \subset \mathcal{X}$

$$\inf_{f \in c_0^+(S)} \frac{\mathrm{Ent}_\pi(f^2)}{\mathrm{Var}_\pi(f)} \geq \frac{\log \frac{1}{\pi(S)}}{1-\pi(S)}\,.$$

Define a probability measure $\pi'(x) = \frac{\pi(x)}{\pi(S)}$ if $x \in S$ and $\pi'(x) = 0$ otherwise. Then

$$\inf_{f \in c_0^+(S)} \frac{\mathrm{Ent}_\pi(f^2)}{\mathrm{Var}_\pi(f)} = \inf_{f \in c_0^+(S)} \frac{\mathrm{Ent}_{\pi'}(f^2) + \log \frac{1}{\pi(S)}\, E_{\pi'} f^2}{E_{\pi'} f^2 - \pi(S)\,(E_{\pi'} f)^2}$$

Rearranging the terms, it suffices to show that

$$\inf_{f \in c_0^+(S)} \frac{\mathrm{Ent}_{\pi'}(f^2)}{\mathrm{Var}_{\pi'}(f)} \geq \frac{\pi(S) \log \frac{1}{\pi(S)}}{1-\pi(S)}\,.$$

However, since $\pi(S) \in (0,1)$ then $\pi(S)\frac{\log(1/\pi(S))}{1-\pi(S)} \leq 1$ and so it suffices that for every probability measure and $f \geq 0$ that $\mathrm{Ent}(f^2)/\mathrm{Var}(f) \geq 1$. This is true, as observed in [LO00] and recalled in Remark 6.7 of [BT03]. $\square$

The bound $\lambda_0(A) \geq \rho \log(1/\pi(A))$ can be shown similarly, but without need for the result of [LO00]. Like log-Sobolev inequalities, Nash inequalities also yield bounds on the spectral profile:

Lemma 3.8. *Given a Nash inequality*

$$\|f\|_2^{2+1/D} \leq C \left[\mathcal{E}(f,f) + \frac{1}{T}\|f\|_2^2\right] \|f\|_1^{1/D}$$

which holds for every function $f : \mathcal{X} \mapsto \mathbb{R}$ and some constants C, D, $T \in \mathbb{R}_+$, then

$$\Lambda(r) \geq \frac{1}{C\, r^{1/2D}} - \frac{1}{T}.$$

Proof. The Nash inequality can be rewritten as

$$\frac{\mathcal{E}(f,f)}{\|f\|_2^2} \geq \frac{1}{C}\left(\frac{\|f\|_2}{\|f\|_1}\right)^{1/D} - \frac{1}{T}$$

Then,

$$\begin{aligned}\lambda_0(A) = \inf_{f \in c_0(A)} \frac{\mathcal{E}(f,f)}{\|f\|_2^2} &\geq \inf_{f \in c_0(A)} \frac{1}{C} \left(\frac{\|f\|_2}{\|f\|_1} \right)^{1/D} - \frac{1}{T} \\ &\geq \frac{1}{C\,\pi(A)^{1/2D}} - \frac{1}{T}\,.\end{aligned}$$

The final inequality was due to Cauchy-Schwartz: $\|f\|_1 \leq \|f\|_2 \sqrt{\pi(\operatorname{supp} f)}$. The lemma follows by minimizing over $\pi(A) \leq r$. $\square$

Although the spectral profile $\Lambda(r)$ is controlled by the spectral gap λ_1 for $r \geq 1/2$, Nash inequalities tend to be better for r close to 0, and log-Sobolev inequalities for intermediate r. Combining Lemmas 3.7 and 3.8, we get the following bounds on mixing time:

Corollary 3.3. *Given the spectral gap λ_1 and the log-Sobolev constant ρ and/or a Nash inequality with $DC \geq T$, $D \geq 1$ and $\pi_* \leq 1/4e$, the L^∞ mixing time for the continuous-time Markov chain with $\epsilon \leq 8$ satisfies*

$$\begin{aligned}\tau_\infty(\epsilon) &\leq \frac{2}{\rho} \log\log \frac{1}{4\pi_*} + \frac{2}{\lambda_1} \log \frac{8}{\epsilon} \\ \tau_\infty(\epsilon) &\leq 4T + \frac{2}{\lambda_1} \left(2D \log \frac{2DC}{T} + \log \frac{4}{\epsilon} \right) \\ \tau_\infty(\epsilon) &\leq 4T + \frac{2}{\rho} \log\log \left(\frac{2DC}{T} \right)^{2D} + \frac{2}{\lambda_1} \log \frac{8}{\epsilon}\end{aligned}$$

Proof. For the first upper bound use the log-Sobolev bound $\Lambda(r) \geq \rho \log(1/r)$ when $r < 1/2$ and the spectral gap bound when $r \geq 1/2$. Simple integration gives the result.

For the second upper bound use the Nash bound when $r \leq (T/2DC)^{2D}$ and spectral gap bound for the remainder. Then

$$\begin{aligned}\tau_\infty(\epsilon) &\leq \int_{4\pi_*}^{(T/2DC)^{2D}} \frac{2dr}{r \, \frac{1}{C\,r^{1/2D}} \left(1 - \frac{C\,r^{1/2D}}{T}\right)} + \int_{(T/2DC)^{2D}}^{4/\epsilon} \frac{2dr}{r\,\lambda_1} \\ &\leq 4T + \frac{2}{\lambda_1} \log \frac{4/\epsilon}{(T/2DC)^{2D}}\end{aligned}$$

where the second inequality used the bound $1 - \frac{C r^{1/2D}}{T} \geq 1 - \frac{1}{2D} \geq 1/2$ before integrating. Simplification gives the result.

For the mixed bound use the Nash bound when $r \leq (T/2DC)^{2D}$, the log-Sobolev bound for $(T/2DC)^{2D} \leq r < 1/2$ and the spectral gap bound when $r \geq 1/2$. □

Similar discrete time bounds follow from Corollary 3.1. When $\forall x : K(x,x) \geq \alpha$ then these bounds are roughly a factor α^{-1} larger than the continuous time case.

These bounds compare well with previous results shown through different methods. For instance, Aldous and Fill [AF] combine results of Diaconis and Saloff-Coste [DSC96a, DSC96b] to show the continuous time bound

$$\tau_\infty(\epsilon) \leq 2T + \frac{1}{\rho} \log\log \left(\frac{DC}{T}\right)^D + \frac{1}{\lambda_1} \left(4 + \log(1/\epsilon)\right)$$

whenever $DC \geq T$.

3.4.3 Walks with Moderate Growth

In this section, we describe how estimates on the volume growth of a walk give estimates on the spectral profile $\Lambda(v)$. The treatment given here is analogous to the method of Nash inequalities described in [DSC96b].

Define the Cayley graph of (K, π) to be the undirected graph on the state space $\mathcal{X}$ with edge set $E = \{(x,y) : \pi(x)K(x,y) + \pi(y)K(y,x) > 0\}$. Let $d(x,y)$ be the usual graph distance, and denote the closed ball of radius r around x by $B(x,r) = \{z : d(x,z) \leq r\}$. The volume of $B(x,r)$ is given by $V(x,r) = \sum_{z \in B(x,r)} \pi(z)$.

Definition 3.8. For $A, d \geq 1$, the finite Markov chain (K, π) has (A, d)-moderate growth if

$$V(x,r) \geq \frac{1}{A}\left(\frac{r+1}{\gamma}\right)^d \quad \forall x \in \mathcal{X},\ 0 \leq r \leq \gamma \tag{3.8}$$

where γ is the diameter of the graph.

For any f and $r \geq 0$, set

$$f_r(x) = \frac{1}{V(x,r)} \sum_{y \in B(x,r)} f(y)\pi(y).$$

Definition 3.9. The finite Markov chain (K, π) satisfies a local Poincaré inequality with constant a if for all f and $r \geq 0$

$$\|f - f_r\|_2^2 \leq ar^2 \mathcal{E}(f,f). \tag{3.9}$$

Under assumptions (3.8) and (3.9), Diaconis and Saloff-Coste [DSC96b] derive the Nash inequality

$$\|f\|_2^{2+4/d} \leq C \left[\mathcal{E}(f,f) + \frac{1}{a\gamma^2}\|f\|_2^2\right] \|f\|_1^{4/d} \tag{3.10}$$

where $C = (1+1/d)^2(1+d)^{2/d}A^{2/d}a\gamma^2$. By Lemma 3.8, this immediately implies the lower bound on the spectral profile

$$\Lambda(v) \geq \left(\frac{d^2}{(d+1)^{2+2/d}A^{2/d}v^{2/d}} - 1\right)\frac{1}{a\gamma^2}.$$

Theorem 3.4 below shows how to bound $\Lambda(r)$ in terms of a local Poincaré inequality and the volume growth function

$$V_*(r) = \inf_x V(x,r).$$

The proof is similar to the derivation of Nash inequalities for walks with moderate growth shown in [DSC96b].

Theorem 3.4. *Let (K, π) be a finite Markov chain that satisfies the local Poincaré inequality with constant a. For $v \leq 1/2$, the spectral profile satisfies*

$$\Lambda(v) \geq \frac{1}{4aW^2(2v)}$$

where $W(v) = \inf\{r : V_*(r) \geq v\}$.

Proof. Fix $S \subset \mathcal{X}$ with $\pi(S) \leq 1/2$ and $f \in c_0(S)$. It is sufficient to show that

$$\frac{\mathcal{E}(f,f)}{\|f\|_2^2} \geq \frac{1}{4aW^2(2\pi(S))}.$$

First observe that

$$\begin{aligned} \|f\|_2^2 &= \langle f - f_r, f\rangle + \langle f_r, f\rangle \\ &\leq \|f - f_r\|_2 \cdot \|f\|_2 + \langle f_r, f\rangle. \end{aligned}$$

Now,

$$\begin{aligned} \langle f_r, f\rangle &= \sum_x \left(\frac{1}{V(x,r)} \sum_{y \in B(x,r)} f(y)\pi(y) \right) f(x)\pi(x) \\ &\leq \frac{1}{V_*(r)} \|f\|_1^2 \\ &\leq \frac{\pi(S)}{V_*(r)} \|f\|_2^2. \end{aligned}$$

Consequently, by the local Poincaré inequality,

$$\|f\|_2^2 \leq \sqrt{a} r \mathcal{E}(f,f)^{1/2} \|f\|_2 + \frac{\pi(S)}{V_*(r)} \|f\|_2^2.$$

Dividing by $\|f\|_2^2$ and choosing $r = W(2\pi(S))$ we have

$$1 \leq \sqrt{a} r \frac{\mathcal{E}(f,f)^{1/2}}{\|f\|_2} + 1/2$$

and the result follows. □

Corollary 3.4. *Let (K, π) be a finite Markov chain that satisfies (A, d)-moderate growth and the local Poincaré inequality with constant a. Then the L^∞ mixing time for the continuous-time chain satisfies*

$$\tau_\infty(\epsilon) \leq C(a, A, d, \epsilon)\gamma^2$$

where γ is the diameter of the graph and $C(a, A, d, \epsilon)$ is a constant depending only on a, A, d and ϵ.

Proof. By the moderate growth assumption, $W(v) \leq \gamma(Av)^{1/d}$. And so for $v \leq 1/2$

$$\Lambda(v) \geq \frac{1}{4aW^2(2v)} \geq \frac{1}{8aA^{2/d}\gamma^2 v^{2/d}}.$$

For $v \geq 1/2$, note that $\Lambda(v) \geq \lambda_1 \geq \Lambda(1/2)/2$. The result now follows immediately from Theorem 1.2. $\square$

In Theorem 3.1 of [DSC94] Diaconis and Saloff-Coste show that for walks on groups with (A, d)-moderate growth, local Poincaré inequality with constant a, and $\gamma \geq A4^{d+1}$

$$\tau_\infty(1/e) \geq \frac{\gamma^2}{4^{2d+1}A^2}.$$

It follows that $\tau_\infty(1/e) = \Theta(\gamma^2)$, and Corollary 3.4 was of the correct order γ^2.

For instance, consider the example of simple random walk on the n-cycle discussed in Section 3.4.1. For this walk $V(x_i, r) = (1+2\lfloor r \rfloor)/n$, and so it satisfies the moderate growth criterion (3.8) with $A = 6$, $d = 1$ and diameter $\gamma = \lfloor n/2 \rfloor$. Moreover, it is shown in [DSC96b] that every group walk satisfies the local Poincaré inequality

$$\|f - f_r\|_2^2 \leq 2|S|r^2\mathcal{E}(f, f)$$

where S is a symmetric generating set for the walk. Consequently, Corollary 3.4 shows that the walk on the n-cycle mixes in $O(n^2)$ time. For several additional examples of walks with moderate growth, see [DSC94, DSC96b].

3.4.4 The Viscek Graphs

For a random walk (K, π) consider its Cayley graph defined in Section 3.4.3. Define the minimum volume of a disk of radius r by $V_*(r) = \min_x\{V(x, r)\}$. Here we first use a result of [BCG01] that shows that the spectral profile $\Lambda(r)$ can be bounded in terms of the volume growth $V_*(r)$ alone (see Lemma 3.9). We then apply this technique to analyze walks on the fractal-like Viscek family of finite graphs.

Lemma 3.9 ([BCG01]). *Let* $Q_* = \inf_{x\sim y}[\pi(x)K(x,y) + \pi(y)K(y,x)]$ *and*

$$w(r) = \inf\{k : V_*(k) > r\}.$$

Then

$$\lambda(A) \geq \frac{Q_*}{2\pi(A)w(\pi(A))}.$$

Proof. Fix $f \in c_0(A)$ normalized so that $\|f\|_\infty = 1$. Then,

$$\|f\|_2^2 = \sum_x |f(x)|^2\pi(x) \leq \pi(A).$$

Let x_0 be a point such that $|f(x_0)| = 1$ and let $k = \max\{l \in \mathbb{N} : B(x_0, l) \subset A\}$. Then there is a sequence of points $x_0, x_1, \ldots, x_{k+1}$ with $x_i \sim x_{i+1}$, $x_0, \ldots, x_k \in A$ and $x_{k+1} \notin A$. So,

$$\begin{aligned}
\mathcal{E}(f,f) &= \frac{1}{2}\sum_{x,y}|f(x)-f(y)|^2\pi(x)K(x,y) \\
&\geq \frac{1}{2}\sum_{i=0}^{k}|f(x_{i+1})-f(x_i)|^2[\pi(x_i)K(x_i,x_{i+1}) + \pi(x_{i+1})K(x_{i+1},x_i)] \\
&\geq \frac{Q_*}{2(k+1)}\left(\sum_{i=0}^{k}|f(x_{i+1})-f(x_i)|\right)^2 \\
&= \frac{Q_*}{2(k+1)}|f(x_{k+1})-f(x_0)|^2 \\
&= \frac{Q_*}{2(k+1)}.
\end{aligned}$$

Consequently,

$$\lambda_0(A) = \inf_{f\in c_0(A)}\frac{\mathcal{E}(f,f)}{\|f\|_2^2} \geq \frac{Q_*}{2(k+1)\pi(A)}.$$

To finish the proof, observe that $\pi(A) \geq V(x_0,k) \geq V_*(k)$, and so $w(\pi(A)) \geq k+1$. ∎

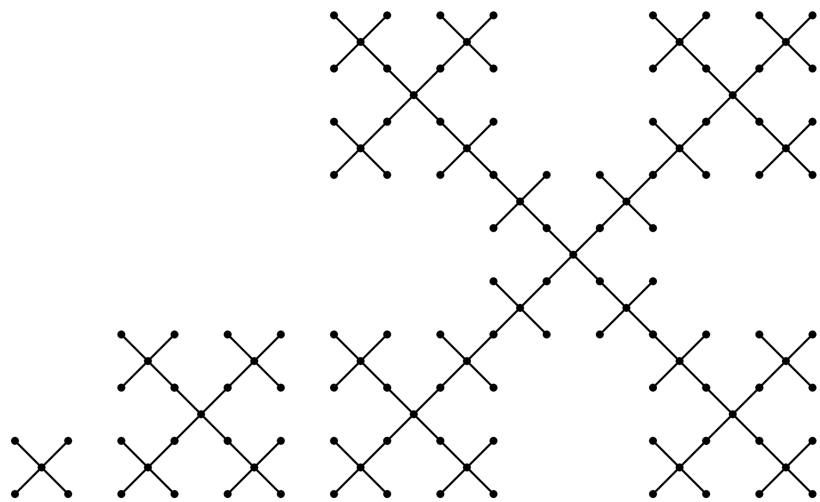

Figure 3.1: The first three generations $\mathcal{V}_4(0)$, $\mathcal{V}_4(1)$ and $\mathcal{V}_4(2)$ of a Viscek graph with $N = 4$.

The Viscek graphs are a two parameter family of finite trees that are inductively defined as follows. Fix the parameter $N \geq 2$, and define $\mathcal{V}_N(0)$ to be the star graph on $N+1$ vertices (i.e. a central vertex surrounded by N vertices). Given $\mathcal{V}_N(n-1)$ choose N vertices $x_1, \ldots, x_N$ such that $d(x_i, x_j) = \operatorname{diam}(\mathcal{V}_N(n-1))$ for $i \neq j$. Construct $\mathcal{V}_N(n)$ by taking $N+1$ copies $\{\mathcal{V}_N^i(n-1)\}_{i=0}^N$ of the $(n-1)^{th}$ generation graph, and for $1 \leq i \leq N$ identifying $x_i^0 \in \mathcal{V}_N^0(n-1)$ with $x_i^i \in \mathcal{V}_N^i(n-1)$. Observe that a different choice of vertices $x_1, \ldots, x_N$ leads to an isomorphic construction. For $N = 2$, $\mathcal{V}_2(n)$ is a path for each n. Figure 3.1 illustrates the first three generations of a Viscek graph for $N = 4$.

The following lemma bounds the spectral profile and mixing time for simple random walk on $\mathcal{V}_N(n)$. The proof is analogous to the volume growth computation for the infinite Viscek graph $\mathcal{V}_N(\infty) = \lim_{n\to\infty} \mathcal{V}_N(n)$ given in [BCG01] and recalled in [PSC].

Lemma 3.10. *For $N \geq 2$, $r \leq 1$ the spectral profile $\Lambda(r)$ for simple random walk on $\mathcal{V}_N(n)$ satisfies*

$$\frac{a(N)}{\gamma^{d+1}r^{1+1/d}} \leq \Lambda(r) \leq \frac{A(N)}{\gamma^{d+1}r^{1+1/d}} \qquad d = \log_3(N+1)$$

where $\gamma = diam(\mathcal{V}_N(n)) = 2 \cdot 3^n$ and the constants $a, A > 0$ depend only on N.

In particular, there exist constants $b, B > 0$ depending only on N such that the mixing time for the continuous-time walk satisfies

$$b(N)\gamma^{d+1} \leq \tau_1(1/e) \leq \tau_\infty(1/e) \leq B(N)\gamma^{d+1}.$$

Observe that since the conductance profile for $\mathcal{V}_N(n)$ satisfies

$$\Phi(r) \approx \frac{1}{|E_N(n)|r} \approx \frac{1}{\gamma^d r},$$

using the conductance profile bound of Corollary 1.2 results in the upper bound $\tau_\infty(1/e) \preceq \gamma^{2d}$ which overestimates the mixing time for $N \geq 3$.

Proof. We first show that the mixing time bound follows from the spectral profile estimate. The upper bound is a direct consequence of Theorem 1.2. Recall that for an ergodic chain, the spectral gap λ_1 and L^1 mixing time are related by $1/\lambda_1 \leq \tau_1(1/e)$ (see e.g. [SC96]). Since $\Lambda(r) \geq \lambda_1$, the lower bound is immediate.

To estimate the spectral profile, first note that the number of edges $|E_N(n)| = N(N+1)^n$. Since $\mathcal{V}_N(n)$ is a tree, $|\mathcal{V}_N(n)| = N(N+1)^n + 1$. Furthermore, $\text{diam}(\mathcal{V}_N(n)) = 2 \cdot 3^n$.

For $0 \leq k \leq n$, define a k-block to be a subgraph of $\mathcal{V}_N(n)$ isomorphic to the k^{th} generation graph $\mathcal{V}_N(k)$. Fix $x \in \mathcal{V}_N(n)$ and $3 \leq r \leq \text{diam}(\mathcal{V}_N(n))$. Then there is a unique integer m such that $3^{m+1} \leq r < 3^{m+2}$. Moreover, the vertex x is contained in some m block B. Since $\text{diam}(B) = 2 \cdot 3^m$, $B(x,r) \supseteq B$. Consequently,

$$|B(x,r)| \geq |B| = N(N+1)^m + 1$$

and since $\pi_* = 1/(2|E_N(n)|)$

$$V_*(r) \geq \frac{N(N+1)^m+1}{2N(N+1)^n} \succeq \left(\frac{r}{\gamma}\right)^d$$

where $d = \log_3(N+1)$ and the notation $a \preceq b$ indicates that there is some constant $c(N) > 0$ depending only on N such that $a \leq c(N)b$. Thus, using the notation of Lemma 3.9, $w(s) \preceq \gamma s^{1/d}$. Since $Q_* = 1/|E_N(n)| \succeq 1/\gamma^d$, Lemma 3.9 gives the lower bound on the spectral profile.

For the upper bound we construct test functions f_m supported on m-blocks. Given an m-block $A \subset \mathcal{V}_N(n)$, choose vertices $x_1, \ldots, x_N$ such that $d(x_i, x_j) = \operatorname{diam}(A)$ for $i \neq j$, and call the shortest paths between these vertices diagonals. These diagonals meet in a unique point o at the center of the m-block. Define the function $f_m \in c_0(A)$ as follows: Along diagonals, f_m varies linearly with $f_m(o) = 1$ and $f_m(x_i) = 0$. Since $d(o, x_i) = \operatorname{diam}(A)/2 = 3^m$, along diagonals the function is given explicitly by $f_m(x) = 1 - d(o,x)/3^m$. For a point x off of the diagonals, let $f_m(x) = f_m(x')$ where x' is the closest point to x that lies on a diagonal. (See Figure 3.2 for a graphical representation of f_m). Now, since $K(x,y)\pi(x) = 1/(2|E_N(n)|)$ for $x \sim y$

$$\begin{aligned}
\mathcal{E}(f_m, f_m) &= \frac{1}{2}\sum_{x,y} |f_m(x) - f_m(y)|^2 K(x,y)\pi(x) \\
&= 3^{-2m} \cdot \frac{N3^m}{2|E_N(n)|} \\
&\approx \frac{1}{\gamma^d 3^m}.
\end{aligned}$$

Define the central $m-1$ block of A to be $A' = \{x \in A : d(o,x) \leq 3^{m-1}\}$. Since $f_m(x) \geq 2/3$ on A',

$$\|f_m\|_2^2 \geq \frac{4}{9}\pi(A') \approx \frac{(N+1)^m}{\gamma^d}.$$

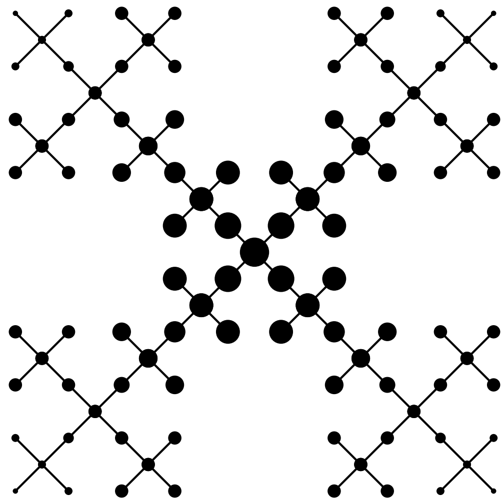

Figure 3.2: A graphical representation of the test function f_2 supported on a 2-block of $\mathcal{V}_4(n)$.

It is sufficient to prove the upper bound on $\Lambda(r)$ for $1/(N+1)^{n-2} < r \leq 1/2$. For these r take

$$m(r) = \left\lfloor \frac{\log r(N+1)^{n-2}}{\log N+1} \right\rfloor \leq n.$$

Then $(N+1)^{m(r)} \leq r(N+1)^{n-2}$ and so for an $m(r)$-block K, $\pi(K) \leq r$. Consequently, for r in this range

$$\begin{aligned} \frac{\mathcal{E}(f_m, f_m)}{\mathrm{Var}(f_m)} &\leq \frac{2\mathcal{E}(f_m, f_m)}{\|f_m\|_2^2} \\ &\preceq \frac{1}{[3(N+1)]^m} \end{aligned}$$

Finally, since $(N+1)^m \succeq r\gamma^d$

$$\begin{aligned} \frac{\mathcal{E}(f_m, f_m)}{\mathrm{Var}(f_m)} &\preceq (N+1)^{-m\frac{\log 3(N+1)}{\log N+1}} \\ &= (N+1)^{-m(1+1/d)} \\ &\preceq \frac{1}{\gamma^{d+1} r^{1+1/d}} \end{aligned}$$

and the upper bound on $\Lambda(r)$ follows. $\square$

3.4.5 The Torus

Consider simple random walk on the product group $\mathbb{Z}_n \times \mathbb{Z}_{n^2}$. For this model, it is not hard to see that $\gamma = \Theta(n^2)$ and that the volume satisfies

$$V_*(r) \asymp \begin{cases} (r+1)^2/n^3 & 0 \le r \le n \\ r/n^2 & n \le r \le n^2 \end{cases}.$$

Taking $r = 0$ in (3.8) shows that walks with moderate growth must have

$$\frac{1}{n^3} \ge \frac{1}{A}\left(\frac{1}{n^2}\right)^d.$$

Consequently, $\mathbb{Z}_n \times \mathbb{Z}_{n^2}$ is of moderate growth with $d = 3/2$ and furthermore, this is the optimal choice of d (assuming A and d are constant). Corollary 3.4 gives the correct $\gamma^2 = n^4$ mixing time, but gives the underestimate

$$\Lambda(v) \ge \frac{C(a, A)}{\gamma^2 v^{4/3}}$$

for the spectral profile. The problem is that the moderate growth criterion alone is not sufficient to identify the two different scales of volume growth present in this example: For $r \ll 1/n$ the space appears 2-dimensional, while for $r \gg 1/n$ it looks 1-dimensional. However, we can apply Theorem 3.4 to directly take into account volume estimates, leading to sharp bounds on both the spectral profile and the rate of decay of $d_{\pi,\infty}(H_t, \pi)$.

Lemma 3.11. *For $a, b \ge 2$, the walk on $G = \mathbb{Z}_a \times \mathbb{Z}_b$ with generating set $\{(\pm 1, 0), (0, \pm 1)\}$ has spectral profile satisfying*

$$\Lambda(v) \asymp \begin{cases} 1/vab & 1/ab \le v \le a/b \\ 1/v^2b^2 & a/b \le v \le 1 \\ 1/b^2 & 1 \le v \end{cases}.$$

In particular,

$$d_{\pi,\infty}(H_t,\pi) \asymp \begin{cases} ab/(t+1) & 0 \le t \le a^2 \\ b/t^{1/2} & a^2 \le t \le b^2 \end{cases}$$

and there are constants $c_1, c_2 > 0$ *such that for* $t \ge b^2$

$$e^{-c_1 t/b^2} \preceq d_{\pi,\infty}(H_t,\pi) \preceq e^{-c_2 t/b^2}.$$

Proof. Without loss of generality, assume $a \le b$. Then $\mathrm{diam}(G) = \Theta(b)$, and the volume function is given by

$$V_*(r) \asymp \begin{cases} (r+1)^2/(ab) & 0 \le r \le a \\ r/b & a \le r \le \mathrm{diam}(G) \end{cases}.$$

Consequently, $W(v) = \inf\{r : V_*(r) \ge v\}$ satisfies

$$W(v) \asymp \begin{cases} v^{1/2}(ab)^{1/2} & 1/ab \le v \le a/b \\ vb & a/b \le v \le 1 \end{cases}.$$

Since this group walk is driven by a constant number of generators, the chain satisfies a local Poincaré inequality with constant independent of a and b. For $v \le 1/2$, the lower bound on the $\Lambda(v)$ now follows from Theorem 3.4. By Lemma 3.2, $\lambda_1 \ge \Lambda(1/2)/2$. For $v \ge 1/2$, the lower bound on $\Lambda(r)$ then follows from the fact $\Lambda(r) \ge \lambda_1$.

For the upper bound, for $m \ge 1$ define linear functions $f_m \in c_0(B(0,m))$ by

$$f_m(x) = 1 - \frac{d_G(0,x)}{m}.$$

Then,

$$\begin{aligned} \mathcal{E}(f_m, f_m) &= \frac{1}{2|G|}\sum_{x,y}[f_m(x) - f_m(y)]^2 K(x,y) \\ &\le \frac{V_*(m)}{2m^2}. \end{aligned}$$

Since G is volume doubling (i.e. $V_*(2m) \approx V_*(m)$)

$$\|f_m\|_2^2 \geq \left(\frac{1}{2}\right)^2 V_*(m/2) \succeq V_*(m)$$

and consequently, for $V_*(m) \leq 1/2$

$$\frac{\mathcal{E}(f_m, f_m)}{\mathrm{Var}(f_m)} \leq 2\frac{\mathcal{E}(f_m, f_m)}{\|f_m\|_2^2} \preceq \frac{1}{m^2}.$$

Since it is sufficient to consider only $v \leq 1/2$, the upper bound on $\Lambda(v)$ follows by taking $m = W(v) - 1$.

Following the notation of Theorem 3.3, define $\gamma(t)$ implicitly through

$$L(r) = \begin{cases} C/rab & 1/ab \leq r \leq a/b \\ C/r^2b^2 & a/b \leq r \leq 1 \end{cases}$$

and observe that for C sufficiently large, the walk satisfies the anti-Faber-Krahn inequality with $L(r)$ by the upper bound on $\Lambda(r)$. Then $\gamma(t)$ is given explicitly as

$$\gamma(t) = \begin{cases} (Ct+1)/ab & 0 \leq t \leq (a^2-1)/C \\ (2Ct - a^2 + 2)^{1/2}/b & (a^2-1)/C \leq t \leq (b^2 + a^2 - 2)/(2C) \end{cases}.$$

Furthermore,

$$\frac{\gamma'(t)}{\gamma(t)} = \begin{cases} C/(Ct+1) & 0 < t \leq (a^2-1)/C \\ C/(2Ct - a^2 + 2) & (a^2-1)/C \leq t < (b^2 + a^2 - 2)/(2C) \end{cases}.$$

So, $\gamma(t)$ is δ-regular on $(0, T)$ with $\delta = 1/6$ and $T = (b^2 + a^2 - 2)/(2C)$. For $t \leq cb^2$ and c sufficiently small, the lower bound on convergence now follows from Theorem 3.3.

Let $\{\lambda_i\}$ be the eigenvalues of $I - K$ with corresponding real orthonormal eigenfunctions $\{\psi_i\}$. Since $h_t(x,y) = H_t 1_y(x)$, writing $1_y(\cdot)$ in this L^2 basis we have

$$h_t(x,y) = \sum_{i=0}^{n-1} e^{-t\lambda_i}\psi_i(x)\psi_i(y) = 1 + \sum_{i=1}^{n-1} e^{-t\lambda_i}\psi_i(x)\psi_i(y).$$

In particular,

$$\sup_x h_t(x,x) - 1 \geq \sup_x e^{-t\lambda_1}\psi_i^2(x) \geq e^{-t\lambda_1}$$

since the eigenfunctions are normalized in L^2. Since $\lambda_1 \approx \Lambda(1/2) \approx 1/b^2$, the lower bound on convergence rate follows.

Now we show the upper bound. By the lower bound on $\Lambda(r)$, using the notation of Theorem 3.1,

$$V(t) \succeq \begin{cases} t/ab & 0 \leq t \leq a^2 \\ t^{1/2}/b & a^2 \leq t \leq b^2 \\ e^{ct/b^2} & b^2 \leq t \end{cases}.$$

Consequently,

$$\sup_{x\in\mathcal{X}} d_{2,\pi}^2(H_t(x,\cdot),\pi) \preceq \begin{cases} ab/t & 0 \leq t \leq a^2 \\ b/t^{1/2} & a^2 \leq t \leq b^2 \\ e^{-ct/b^2} & b^2 \leq t \end{cases}.$$

Since the walk is reversible, the upper bound on $d_{\infty,\pi}$ then follows from the argument of Theorem 1.2 $\square$

For $a \ll b$, Lemma 3.11 captures the fact that decay is fast at the start of the walk and then slows down. More specifically,

$$d_{\pi,\infty}(H_t,\pi) \asymp 1/V_*(t^{1/2}) \quad t \leq \operatorname{diam}^2(G).$$

As shown in [DSC95b], this relationship holds in general for random walks on groups with volume doubling. While we considered only the simple case of $\mathbb{Z}_a \times \mathbb{Z}_b$, the technique applies well to more general k-fold products.

Chapter 4

Analysis of Top to Bottom Shuffles

4.1 Background

A deck of n cards can be shuffled by repeatedly removing the top card and inserting it uniformly at random back into the deck. A coupling argument shows that the total variation mixing time for this Markov chain is $n \log n$ (see e.g. [Ald82, Dia88, SC04a]). In fact, a detailed analysis yields a closed form expression for the distribution of this chain after any number of steps (see [DFP92]).

Here we analyze a class of walks that generalizes the top to random chain, namely top to bottom shuffles. These shuffles are generated by moving the top card uniformly at random to any of the bottom k_n positions of the deck. For $k_n = n$ we recover the top to random walk. For $k_n = 2$, this is the Rudvalis shuffle, and upper and lower bounds of order $n^3 \log n$ have been shown by Hildebrand [Hil90] and Wilson [Wil03b] respectively.

More formally, let S_n be the permutation group, and let $\sigma \in S_n$ denote an element of this group, interpreting $\sigma(i) = j$ to mean that position i holds the card with label j. Fix $n \geq k_n > 1$, and denote a cycle permutation by

$$\sigma_l = (1, 2, \ldots, l)$$

where $\sigma_l(i) = i + 1$ for $1 \leq i \leq l - 1$, $\sigma_l(l) = 1$, and $\sigma_l(i) = i$ otherwise. Define the probability measure q_{n,k_n} on S_n by

$$q_{n,k_n}(\sigma) = \begin{cases} \frac{1}{k_n} & \text{if } \sigma = \sigma_l \text{ for some } n - k_n + 1 \leq l \leq n \\ 0 & \text{otherwise} \end{cases}$$

and let π be the uniform distribution on S_n. Then the top to bottom shuffle driven by q_{n,k_n} is non-reversible, aperiodic and irreducible with stationary distribution π.

Let q^*_{n,k_n} denote the bottom k_n to top shuffle. It is well known that studying this shuffle is equivalent to studying q_{n,k_n}. Then for the top to bottom k_n walk q_{n,k_n}, and the reversible variant $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$, we derive bounds on the total variation and L^2 mixing times T_1 and T_2. In particular, our main results are summarized below, and can also be found in [Goe]. In these statements, $A(c)$, $B(c)$, etc., denote positive, finite constants that may depend on the fixed parameter c but not on n.

Theorem 1.3. For the top to bottom k_n shuffles q_{n,k_n}

1. if $k_n \geq n - \sqrt{(n \log n)/2}$ then
$$T_1 \sim n \log n.$$
That is, the walk presents a total variation cut-off at time $n \log n$.

2. if $k_n \geq cn$ with $c \in (0, 1)$ then
$$A(c)n \log n \leq T_1 \leq B(c)n^2 \log^2 n$$

3. if $k_n \leq C$ then
$$A(C)n^3 \leq T_1 \leq B(C)n^3 \log n$$

4. if $k_n = 2, 3$ then
$$An^3 \log n \leq T_1 \leq Bn^3 \log n.$$

For Theorem 1.3(1), the walk presents a total variation cut-off at time $n \log n$. See Lemma 4.2 for a precise statement of the result. Using techniques similar to those presented here, Jonasson [Jon] has recently shown that top to bottom shuffles have total variation mixing time $T_1 \approx n^3 \log n/k_n^2$.

Theorem 1.4. For the reversible top to bottom k_n shuffles $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$

1. if $k_n \geq n - C$ then
$$T_1 \leq T_2 \leq B(C)n \log n$$
2. if $k_n \leq cn$ with $c \in (0,1)$ then
$$T_2 \geq T_1 \geq A(c)n^2 \quad \text{and} \quad T_2 \geq \frac{A(c)n^3}{k_n^2} \log n$$
3. for any k_n
$$T_1 \leq T_2 \leq Bn^3 \log n.$$

In particular, if $k_n \leq C$ then
$$A(C)n^3 \log n \leq T_2 \leq Bn^3 \log n.$$

The two lower bounds in Theorem 1.4(2) are complementary in the sense that the first gives better estimates for $k_n \approx cn$, while the second works best for $k_n \ll cn$.

Finally, we show that Theorem 1.3 and Theorem 1.4 yield bounds on the lazy top to bottom k_n shuffle
$$\hat{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + \delta_e)$$
where we put weight on the identity.

Theorem 4.1. *For the lazy top to bottom k_n shuffle $\hat{q}_{n,k_n}$*

1. *if $k_n \geq n - C$ then*
$$An \log n \leq T_2 \leq B(C)n \log n$$
2. *if $k_n \geq n - \sqrt{(n \log n)/2}$ then*
$$T_1 \sim 2n \log n$$

3. if $k_n \geq cn$ *with* $c \in (0,1)$ *then*

$$A(c)n \log n \leq T_1 \leq B(c)n^2 \log^2 n$$

4. if $k_n = 2, 3$ *then*

$$An^3 \log n \leq T_1 \leq Bn^3 \log n.$$

5. for any k_n

$$T_2 \leq Bn^3 \log n.$$

For Theorem 4.1(2), the walk presents a total variation cut-off at time $2n \log n$. See Remark 4.3 for a precise statement of the result. Also observe that the estimates of Theorem 4.1(1, 2) bound the L^2 mixing time T_2 and the total variation mixing time T_1 respectively.

As k_n varies from a constant to n, these results are most satisfactory at the extremes of the range. For large k_n the walks behave like the top to random chain, mixing in $n \log n$ steps. Theorem 1.3(1) proves mixing in the strongest possible sense: cut-off at precisely $n \log n$. Let us note here that the precise L^2 cut-off time is not yet known even for the top to random shuffle $q_{n,n}$. For small k_n the walks behave like the Rudvalis shuffle, mixing in $n^3 \log n$ steps. Theorem 1.4 proves this for the reversible chain, whereas Theorem 1.3 and Theorem 4.1 give complete results only for $k_n = 2, 3$.

The worst gap in these results occurs when $k_n \approx n/2$. For these “top to bottom half” shuffles, [Jon] shows a $\Theta(n \log n)$ mixing time for the non-reversible shuffle, and our results give an $\Omega(n^2)$ lower bound for the reversible shuffle. In particular, the non-reversible and reversible top to bottom half shuffles mix at different rates. In this range, one difficulty in analyzing the reversible walk is that comparison with random transposition, one of the best understood models of random walk, can at best yield $O(n^3 \log n)$ upper bounds (see Lemma 4.10).

A variety of methods are used to prove the results of this chapter. The upper bounds for the non-reversible top to bottom shuffle are found by coupling arguments. The lower bound in Theorem 1.3(4) uses Wilson's lemma (see e.g. [SC04b, Wil03b]). For the reversible chain, we use comparison techniques for walks on finite groups to prove both upper and lower bounds (see e.g. [DSC93b]). Notably, comparison previously has been applied only to find upper bounds. It appears that this is the first application of comparison techniques to prove lower bounds.

Sections 4.2 and 4.3 give proofs of Theorem 1.3 and Theorem 1.4 respectively. Section 4.4 applies these results to find bounds on the lazy walk $\hat{q}_{n,k_n}$.

4.2 The Non-Reversible Shuffle

In this section we present upper and lower bounds for the mixing time of the non-reversible walk q_{n,k_n}, using primarily probabilistic techniques. For $k_n = 3$ we use the method of [Wil03b] to derive a lower bound.

To prove mixing time bounds for the top to bottom shuffle, we make extensive use of the following well-known coupling result (see e.g. [Ald82, Dia88, SC04a]).

Proposition 4.1. *Let q be a probability measure on a finite group G. Let (X_n^1, X_n^2) be a coupling for the random walk driven by q with (X_n^1) starting at the identity and (X_n^2) stationary. Then for the stationary measure π*

$$\|q^m - \pi\|_{TV} \leq \mathbb{P}(T > m)$$

where

$$T = \inf\{m \mid \forall k \geq m,\ X_k^1 = X_k^2\}.$$

Furthermore, there exists a coupling such that the inequality above is an equality.

We will also make use of the following coupon-collectors lemma (see [Ald82]).

Lemma 4.1 (Coupon-Collectors Lemma). *Let R_m be the number of distinct cards obtained in m uniform random draws with replacement from a deck of n cards. That is $R_m = |\{C_1, \ldots, C_m\}|$ with C_i iid uniform on $\{1, \ldots, n\}$. Let $L_j = \min\{m|\ R_m = n - j\}$, i.e. the number of draws before all but j cards have been chosen. Then for fixed j,*

$$\frac{L_j}{n \log n} \to 1 \qquad \textit{in probability.}$$

In the case of $q_{n,n}$, i.e. the top to random shuffle, the correct mixing time $n \log n$ can be found using a coupling of the time reversed process $q_{n,n}^*$. For this random to top shuffle the coupling is as follows: Choose a label uniformly at random from $\{1, \ldots n\}$ and in each deck move the card with this label to the top. Clearly this is a coupling, and the coupling time is given by the Coupon-Collectors Lemma (for details, see e.g. [Ald82]). The proof of Lemma 4.2 is by a similar coupling.

Lemma 4.2. *For $k_n \geq n - \sqrt{\frac{1}{2} n \log n}$, the walk (S_n, q_{n,k_n}) presents a total variation cut-off at $t_n = n \log n$. That is, for $\epsilon \in (0,1)$*

$$\lim_{n\to\infty} \|q_{n,k_n}^{(1+\epsilon) n \log n} - \pi\|_{TV} = 0$$

and

$$\lim_{n\to\infty} \|q_{n,k_n}^{(1-\epsilon) n \log n} - \pi\|_{TV} = 1.$$

Proof. Since $d_{TV}(p^{(n)}, u) = d_{TV}(p^{*(n)}, u)$, we can consider the reversed random walk q^*_{n,k_n}. For this reversed walk we define a coupling (X_1^m, X_2^m) where X_1 starts from the identity and X_2 is drawn from the stationary distribution. Let

$$A_j^m = \{X_j^m(i) \mid n - k_n < i \leq n\} \qquad j = 1, 2.$$

That is, A_j^m is the set of cards that at time m are in the bottom k_n positions of deck j. At time m, in the first deck pick a card σ_a uniformly at random from A_1^m and move it to the top of the deck. If $\sigma_a \in A_2^m$, then move card σ_a in the second deck to the top. If not, then in the second deck uniformly at random pick a card from $A_2^m \setminus A_1^m$ and move it to the top.

Clearly, deck one is driven by q^*_{n,k_n}. For the second deck, note that any card in $A_1^m \cap A_2^m$ is chosen if and only if it is chosen in the first deck, and hence with probability $1/k_n$. And cards in $A_2^m \setminus A_1^m$ are chosen with probability

$$\frac{k_n - |A_1^m \cap A_2^m|}{k_n} \cdot \frac{1}{k_n - |A_1^m \cap A_2^m|} = \frac{1}{k_n}.$$

So this is in fact a coupling. Define

$$\tau_0 = \inf\{m \mid X_1^m(i) = X_2^m(i) \text{ for } 1 \leq i \leq n - k_n\}.$$

That is, τ_0 is the first time the top $n - k_n$ cards are matched in both decks. Then for $m > \tau_0$, $A_1^m = A_2^m$, i.e. the set of cards in the bottom k_n positions are the same in each deck. Consequently, after time τ_0 new matches are not broken and every time an unmatched card is chosen, a new match is made.

First we estimate τ_0. Let L be the probability that starting with all cards unmatched, $n-k_n$ consecutive matches are made. Then,

$$\begin{aligned} L &\geq \left(1-\frac{n-k_n}{k_n}\right)^{n-k_n} \\ &\geq \left(1-\frac{1}{\frac{\sqrt{n}}{\sqrt{\frac{1}{2}\log n}}-1}\right)^{\sqrt{\frac{1}{2}n\log n}} \\ &\approx \frac{1}{\sqrt{n}}. \end{aligned}$$

Furthermore, by the Markov property, for fixed $\epsilon > 0$

$$\begin{aligned} P(\tau_0 \geq \epsilon n \log n) &\leq P\left(\tau_0 \geq \epsilon\sqrt{n\log n}\cdot\sqrt{\frac{1}{2}n\log n}\right) \\ &\leq \left[1-\left(1-\frac{1}{\frac{\sqrt{n}}{\sqrt{\frac{1}{2}\log n}}-1}\right)^{\sqrt{\frac{1}{2}n\log n}}\right]^{\epsilon\sqrt{n\log n}} \\ &\stackrel{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Let τ_1 be the time it takes after τ_0 for each card in $A_1^{\tau_0} = A_2^{\tau_0}$ to be selected. That is,

$$\tau_1 = \inf\{m \mid m > 0,\ \text{each card in } A_1^{\tau_0} \text{ has been selected by time } m+\tau_0\}.$$

By the Coupon-Collectors Lemma, for fixed $\epsilon > 0$

$$\lim_{n\to\infty} P(\tau_1 \geq (1+\epsilon)k_n \log k_n) = 0.$$

Finally, if T is the coupling time, then since

$$\begin{aligned} P(T > (1+\epsilon)n\log n) &\leq P\left(\tau_0 \geq \frac{\epsilon}{2}n\log n\right) + P\left(\tau_1 \geq \left(1+\frac{\epsilon}{2}\right)n\log n\right) \\ &\stackrel{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

by Proposition 4.1,

$$\lim_{n\to\infty} \|q_{n,k_n}^{(1+\epsilon)n\log n} - \pi\|_{TV} = 0.$$

The lower bound argument is analogous to that of the top to random shuffle (see e.g. [Ald82]). Let B_j be the set of permutations σ for which the bottom j cards have increasing labels. That is,

$$\sigma(n-j+1) < \sigma(n-j+2) < \cdots < \sigma(n).$$

Then $\pi(B_j) = \frac{1}{j!}$. Starting from the identity, Let L_j be the number of shuffles until all but j of the cards with labels in $\{n-k_n+1, \ldots, n\}$ have been chosen. Then if $L_j > m$ the bottom j cards after m bottom k_n to top shuffles are in increasing order. So for fixed $\epsilon > 0$, there is an $\epsilon' > 0$ such that

$$\begin{aligned} \lim_{n\to\infty} \|q_{n,k_n}^{(1-\epsilon)n\log n} - \pi\|_{TV} &\geq \lim_{n\to\infty} P(L_j > (1-\epsilon)n\log n) - \frac{1}{j!} \\ &\geq \lim_{n\to\infty} P(L_j > (1-\epsilon')k_n \log k_n) - \frac{1}{j!} \end{aligned}$$

since

$$\lim_{n\to\infty} \frac{k_n \log k_n}{n \log n} = 1.$$

Using the Coupon-Collectors Lemma, the result follows. □

Lemma 4.3 and Lemma 4.4 below bound the mixing time of q_{n,k_n} in the cases where k_n is relatively large and when k_n is small. Both lemmas rely on the following coupling.

We construct a coupling (X_1^m, X_2^m) where X_1 starts from the identity and X_2 is drawn from the uniform distribution. Recall that the notation $X_s^m(i) = j$ can be interpreted to mean that at time m position i in deck s holds the card with label j. Let

$$A_s^m = \{X_s^m(i) \mid n-k_n+2 \leq i \leq n\} \qquad s = 1, 2.$$

Note that A_s^m is not the set of cards in the bottom k_n positions (to which the top card can be sent), but rather only the cards in the bottom $k_n - 1$ positions.

We define a coupling as follows: First pick one of the two decks with equal probability. Say we picked deck one. Then X_1 proceeds as usual by uniformly at random moving the top card to one of the bottom k_n positions; X_2 mimics the moves of X_1 except in a couple of cases. If $X_1^m(1) \in A_2^m$ (i.e. the top card in the first deck is in A_2^m), and the first deck moves the top card to position $(X_2^m)^{-1}(X_1^m(1))$, then the second deck moves the top card to $(X_2^m)^{-1}(X_1^m(1)) - 1$. And, if $X_1^m(1) \in A_2^m$ and the first deck moves the top card to $(X_2^m)^{-1}(X_1^m(1)) - 1$, then the second deck moves the top card to $(X_2^m)^{-1}(X_1^m(1))$. We have an analogous description if we originally picked deck two. Accordingly, if card i is on the top of one deck and in the bottom $k_n - 1$ positions of the other deck, then with probability $1/k_n$ it will couple on the next move. Furthermore, matches between the decks are never broken.

Lemma 4.3. *For $c \in (0,1)$ and $k_n \geq cn$, there exist constants $A(c)$ such that the total variation mixing time for the walk driven by q_{n,k_n} satisfies*

$$T_1 \leq An^2 \log^2 n.$$

Proof. We use the coupling described above. Let τ_j be the first time that the cards with label j couple in the two decks. That is,

$$\tau_j = \inf\{m \mid (X_1^m)^{-1}(j) = (X_2^m)^{-1}(j)\}.$$

We estimate τ_j by first showing that starting from any permutation of the decks, any card j has probability at least $\frac{C}{n}$ to couple within $3n \log n$ steps. Let τ_σ^j be the first time card j reaches the top of deck one, starting from state σ. And let τ_σ be the first time the bottom card reaches position $n - k_n$. Then for n sufficiently large

$$\begin{aligned}
P\left(\tau_\sigma^j > 2n\log n\right) &\leq P\left(\tau_\sigma > 2n\log n - (n-k_n)\right)\\
&\leq k_n \exp\left(-\frac{2n\log n - (n-k_n)}{k_n}\right)\\
&\leq \frac{1}{2}.
\end{aligned}$$

The second inequality follows from the fact that τ_σ is the sum of independent geometric waiting times with means $k_n, k_n/2, \dots, k_n/k_n$, and consequently is equivalent to the coupon collectors problem. In particular, the above shows that starting from any state, there is positive probability independent of n and k_n that card j reaches the top of the first deck in $2n\log n$ steps.

When card j gets to the top of the first deck, we are in one of three situations: card j is already coupled, card j in the second deck is in the bottom $k_n - 1$ positions, or card j in the second deck is in the top $n - k_n + 1$ positions. In the first two situations, card j will be coupled at the next step with probability at least $1/k_n$ (if j is already coupled, it will remain coupled at the next step). So we only need to consider the third situation. Assume card j moves to one of the bottom $\lceil Bk_n \rceil$ positions for some $B \in (0,1)$ (which happens with probability at least B). Let τ_B be the first time j leaves the bottom $k_n - 1$ positions. Then since τ_B is the sum of independent geometric waiting times,

$$\begin{aligned}
E\tau_B &\geq \sum_{r=\lceil Bk_n\rceil}^{k_n-1} \frac{k_n}{r}\\
&\geq k_n \log\frac{1}{B + 1/k_n}.
\end{aligned}$$

And,

$$\begin{aligned}
\mathrm{Var}(\tau_B) &\leq \sum_{r=1}^{k_n-1} \frac{k_n(k_n - r)}{r^2}\\
&\leq 2k_n^2.
\end{aligned}$$

By Chebyshev's Inequality,

$$\begin{aligned} P\left(\tau_B > \frac{E\tau_B}{2}\right) &\geq 1 - \frac{4\mathrm{Var}(\tau_B)}{(E\tau_B)^2} \\ &\geq 1 - \frac{8}{\log^2 \frac{1}{B+1/k_n}}. \end{aligned}$$

Consequently, if we choose B and K such that

$$\log \frac{1}{B+1/K} \geq \max\left(\frac{2(1-c)}{c}, 3\right)$$

then there exists $\delta > 0$ (independent of n) such that for $k_n \geq K$,

$$P(\tau_B > n - k_n) > \delta.$$

But if $\tau_B > n - k_n$, then j will still be in the bottom $k_n - 1$ positions of deck one when j reaches the top of deck two. Consequently, for each of the original three cases, after reaching the top of deck one, card j couples within the next $n - k_n$ steps with probability at least δ/n. Combining this with the bound on τ_σ^j, for the coupling time τ_j of card j we have

$$P(\tau_j \leq 3n \log n) \geq \frac{\delta}{2n}.$$

Moreover, by the Markov property

$$\begin{aligned} P(\tau_j > An^2 \log^2 n) &\leq \left(1 - \frac{\delta}{2n}\right)^{\frac{An \log n}{3}} \\ &\leq \exp\left(-\frac{\delta A \log n}{6}\right). \end{aligned}$$

Finally, if T is the coupling time for the two decks, then

$$P(T > An^2 \log^2 n) \leq n \exp\left(-\frac{\delta A \log n}{6}\right)$$

and the result follows by taking A sufficiently large. $\square$

Remark 4.1. Using the lower bound argument of Lemma 4.2, we can show that for $c \in (0,1)$, $k_n \geq cn$, there exist constants $B(c)$ such that the mixing time satisfies

$$T_1 \geq B(c) n \log n.$$

The following lemma gives an upper bound on the mixing time for the walk driven by q_{n,k_n} with $k_n \leq C$. The coupling used to prove the result is the same as in Lemma 4.3, however we analyze the coupling time by a different technique.

Lemma 4.4. *For $k_n \leq C$, there exist constants $A(C)$ such that the total variation mixing time for the walk driven by q_{n,k_n} satisfies*

$$T_1 \leq A n^3 \log n.$$

Proof. Using the coupling described above, we show that starting from any permutation of the decks, any card i has probability at least $\delta > 0$ (independent of n) to couple within n^3 steps. Fix card i and let τ be the first time that card i is on the top of one deck and in the bottom $k_n - 1$ positions of the other. Then at the next step, the cards have probability $1/k_n$ to couple. Let

$$\tau_j^1 = \inf\{t \mid X_j^t(1) = i\}$$

$$\tau_j^m = \inf\{t > \tau_j^{m-1} \mid X_j^t(1) = 1\}.$$

That is, τ_j^m is the time when card i is on top of deck j for the m^{th} time. Without loss of generality, assume that $\tau_1^1 \leq \tau_2^1$. If $\tau_j^m \leq \tau$, then

$$\tau_1^m \leq \tau_2^m \leq \tau_1^m + n - k_n.$$

And if $\tau_j^{m+1} \leq \tau$ then

$$\tau_j^{m+1} \leq \tau_j^m + 2(n - k_n).$$

Define the random variables $d_i^m = [(X_1^m)^{-1}(i) - (X_2^m)^{-1}(i)] \bmod n$ which give the oriented distance between the positions of the i^{th} card in each deck. Note that d_i^m only changes when i is in the bottom $k_n - 1$ positions in at least one deck. Let $\tau^* = \inf\{t > \tau_1^1 \, | X_1^t(i) = n - k_n + 1\}$. Then define Y_h^l as iid random variables with distribution given by

$$P(Y_h^l = t) \stackrel{\text{def}}{=} P(\tau^* - \tau_1^1 = t).$$

That is, Y_h^l gives the amount of time it takes a card to get from the top of the deck to the $n - k_n + 1$ position. Furthermore, before τ the distribution of the change in distance is given by $d_i^{\tau_1^{m+1}} - d_i^{\tau_1^m} \stackrel{\text{dist}}{=} Y_1^m - Y_2^m$. Consequently,

$$P(\tau_1^{m+1} \le \tau) \le P\left(\left|\sum_{l=1}^{m} Y_1^l - Y_2^l\right| \le n\right).$$

Let $\sigma^2 = \text{Var}(Y_1^l - Y_2^l)$, and note that $\sigma < \infty$ since Y_h^l can be realized as a finite sum of geometric waiting times. Then by the central limit theorem, by taking $m = n^2$, we have that $P(\tau_1^{n^2+1} \le \tau) \le 1 - \epsilon$ independent of n. That is, $P(\tau < \tau_1^{n^2+1}) > \epsilon$. Furthermore, since $\tau_1^1 \le n$ with positive probability independent of n, $P(\tau < 3n^3) > \epsilon$. Consequently, there is a $\delta > 0$ such that if τ_i is the coupling time for card i then $P(\tau_i < 3n^3) > \delta$. Finally, if T is the coupling time for the two decks, then

$$\begin{aligned} P(T > An^3 \log n) &\le nP(\tau_i > An^3 \log n) \\ &\le n(1-\delta)^{\frac{A \log n}{3}} \\ &\stackrel{A\to\infty}{\longrightarrow} 0. \end{aligned}$$

By taking A sufficiently large, the result follows from Proposition 4.1. □

Remark 4.2. For $k_n \le C$, the walk performed by one card under the measure q_{n,k_n} is an example of a class of walks known as necklace chains. By results in

[Wil03a], this immediately yields the lower bound

$$T_1 \geq B(C)n^3.$$

In [Hil90], the Rudvalis shuffle $q_{n,2}$ is shown to have an upper bound of order $O(n^3 \log n)$. In [Wil03b], a matching lower bound for this shuffle is given by using Proposition 4.2. Here we show that the method of [Wil03b] can also used to lower bound $q_{n,3}$.

Proposition 4.2. *Suppose that a Markov chains X_t has a lifting (X_t, Y_t), and that Ψ is an eigenfunction of the lifted Markov chain: $E[\Psi(X_{t+1}, Y_{t+1})|(X_t, Y_t)] = \lambda\Psi(X_t, Y_t)$. Suppose that $|\Psi(x, y)|$ is a function of x alone, $|\lambda| < 1$, $\Re(\lambda) \geq 1/2$, and that we have an upper bound R on $E[|\Psi(X_{t+1}, Y_{t+1}) - \Psi(X_t, Y_t)|^2 \mid (X_t, Y_t)]$. Let $\gamma = 1 - \Re(\lambda)$. Then when the number of steps t is bound by*

$$t \leq \frac{\log \Psi_{\max} + \frac{1}{2} \log \frac{\gamma\epsilon}{4R}}{-\log(1-\gamma)}$$

the variation distance satisfies $\|X_t - \pi\|_{TV} \geq 1 - \epsilon$.

For a discussion of Proposition 4.2, see [Wil03b, Wil04, SC04b].

Lemma 4.5. *For $\epsilon > 0$, there exist constants $C(\epsilon) > 0$ such that for $m \leq Cn^3 \log n$*

$$\|q_{n,3}^m - \pi\|_{TV} \geq 1 - \epsilon$$

Proof. Let $X_t^{-1}(j) = j'$ denote that the card with label j is at position j' at time t. First we lift the chain to $(X_t^{-1}, Y_t) = (X_t^{-1}, t \bmod n)$. Let $Z_t(j) = (X_t^{-1}(j) - X_0^{-1}(j) + Y_t(j)) \bmod n$ and let $\eta(t) \in \{\sigma_{n-2}, \sigma_{n-1}, \sigma_n\}$ denote the cycle that is chosen at time t. Then,

$$(X_{t+1}^{-1}(j), Z_{t+1}(j)) = \begin{cases} (X_t^{-1}(j), Z_t(j)+1) & \eta_t = \sigma_{n-1}, X_t^{-1}(j) = n \text{ or} \\ & \eta_t = \sigma_{n-2}, X_t^{-1}(j) \geq n-1 \\ (X_t^{-1}(j)-1, Z_t(j)) & \eta_t = \sigma_n \text{ or} \\ & \eta_t = \sigma_{n-1}, X_t^{-1}(j) \leq n-1 \text{ or} \\ & \eta_t = \sigma_{n-2}, X_t^{-1}(j) \leq n-2 \\ (n-1, Z_t(j)-1) & \eta_t = \sigma_{n-1}, X_t^{-1}(j) = 1 \\ (n-2, Z_t(j)-2) & \eta_t = \sigma_{n-2}, X_t^{-1}(j) = 1 \end{cases}$$

Define $v(x)$ to be the x^{th} number in the list

$$\lambda^{n-3}, \ldots, \lambda, 1, \chi_1, \chi_0$$

and define the functions

$$\Psi_j(X_t^{-1}, Y_t) = v(X_t^{-1}(j)) \exp(Z_t(j) 2\pi i/n)$$
$$\Psi(X_t^{-1}, Y_t) = \sum_{j=1}^{n} \Psi_j(X_t^{-1}, Y_t).$$

Now we will find values for λ, χ_1, χ_0 that make Ψ_j (and hence Ψ) an eigenfunction. Also note that $|\Psi(X_t^{-1}, y_1)| = |\Psi(X_t^{-1}, y_2)|$ for all y_1, y_2. If $2 \geq X_t^{-1}(j) \geq n-2$, then

$$\Psi_j(X_{t+1}^{-1}, Y_{t+1}) = \lambda \Psi_i(X_t^{-1}, Y_t).$$

Let $w = e^{2\pi i/n}$. By looking at what happens when $X_t^{-1}(j) = 1$, $X_t^{-1}(j) = n$, and $X_t^{-1}(j) = n-1$ we find that Ψ_j is an eigenfunction with eigenvalue λ when the equations

$$\begin{aligned} \chi_0 + \chi_1 w^{-1} + w^{-2} &= 3\lambda^{n-2}. \\ \frac{\chi_1}{\chi_0} + 2w &= 3\lambda \\ \frac{2}{\chi_1} + w &= 3\lambda \end{aligned}$$

are satisfied. In particular,

$$\begin{aligned}\chi_0 &= \frac{2}{(3\lambda - w)(3\lambda - 2w)}\\ \chi_1 &= \frac{2}{3\lambda - w}\end{aligned}$$

and λ is a root of the polynomial

$$f(\lambda) = 9\lambda^n - 9w\lambda^{n-1} + 2w\lambda^{n-2} - 3w^{-2}\lambda^2 + w^{-1}\lambda.$$

We will use Newton's method to approximate a root of $f(\lambda)$ starting with $z_0 = 1$ and $z_{k+1} = z_k - f(z_k)/f'(z_k)$. By Taylor's theorem,

$$|f(z_{k+1})| \leq \frac{1}{2} \max_{0 \leq p \leq 1} |f''(pz_k + (1-p)z_{k+1})| \cdot \left| \frac{f(z_k)}{f'(z_k)} \right|^2.$$

Furthermore, since

$$\begin{aligned}f' &= 9n\lambda^{n-1} - 9(n-1)w\lambda^{n-2} + 2w(n-2)\lambda^{n-3} - 6w^{-2}\lambda + w^{-1}\\ f'' &= 9n(n-1)\lambda^{n-2} - 9(n-1)(n-2)w\lambda^{n-3} + 2w(n-2)(n-3)\lambda^{n-4} - 6w^{-2}\end{aligned}$$

if $z = 1 + O(1/n^2)$, then $f'(z) = 2n + O(1)$ and $f''(z) = 2n + O(n)$. So if $z_k = 1 + O(1/n^2)$ and $z_{k+1} = 1 + O(1/n^2)$, then

$$|f(z_{k+1})| \leq \frac{1 + O(1/n)}{4} |f(z_k)|^2.$$

Furthermore,

$$\begin{aligned}f(z_0) &= 9 - 7w + w^{-1} - 3w^{-2}\\ &= \frac{36\pi^2}{n^2} - i\frac{4\pi}{n} + O(1/n^4)\end{aligned}$$

Consequently, by induction,

$$\begin{aligned}|f(z_k)| &\leq 4\left(\frac{\pi}{n}\right)^{2^k} + O\left(\frac{1}{n^{2^k+1}}\right)\\ |z_{k+1} - z_k| &= \frac{2}{n}\left(\frac{\pi}{n}\right)^{2^k} + O\left(\frac{1}{n^{2^k+2}}\right).\end{aligned}$$

So for n sufficiently large, the sequence $\{z_k\}$ converges to a point z_∞ and by continuity, $f(z_\infty) = 0$. Furthermore, since

$$\begin{aligned} f'(z_0) &= 9n - 9(n-1)w + 2w(n-2) - 6w^{-2} + w^{-1} \\ &= 2n - i14\pi + O(1/n) \end{aligned}$$

$$\begin{aligned} \mathrm{Re}(z_1) &= 1 - \mathrm{Re}\left(\frac{f(z_0)}{f'(z_0)}\right) \\ &= 1 - \frac{\mathrm{Re}(f(z_0))\mathrm{Re}(f'(z_0)) + \mathrm{Im}(f(z_0))\mathrm{Im}(f'(z_0))}{|f'(z_0)|^2} \\ &= 1 - \left(\frac{18\pi^2 + 14\pi}{n^3}\right) + O(1/n^4). \end{aligned}$$

Finally, since

$$|z_1 - z_\infty| \leq \frac{2\pi^2}{n^3} + O(1/n^4),$$

there exist $c_2 > c_1 > 0$ such that

$$1 - \frac{c_1}{n^3} + O(1/n^4) \geq \mathrm{Re}(z_\infty) \geq 1 - \frac{c_2}{n^3} + O(1/n^4).$$

With $\lambda = z_\infty$, $\chi_0 = 1 + O(1/n)$, and $\chi_1 = 1 + O(1/n)$. Consequently,

$$\Psi_{\max} = n + O(1/n).$$

Now we estimate R. Since $|\lambda - 1| = O(1/n^2)$,

$$\frac{\Psi_i(X_{t+1}^{-1}, Y_{t+1}) - \Psi_i(X_t^{-1}, Y_t)}{w^{Z_t(i)}} = \begin{cases} (\lambda-1)\lambda^{X_t^{-1}(i)} = O(1/n^2) & 2 \leq X_t^{-1}(i) \leq n-2 \\ \chi_0 - \lambda^{n-3} = O(1/n) & X_t(i) = 1, \eta_t = \sigma_n \\ \chi_1 w^{-1} - \lambda^{n-3} = O(1/n) & X_t(i) = 1, \eta_t = \sigma_{n-1} \\ w^{-2} - \lambda^{n-3} = O(1/n) & X_t(i) = 1, \eta_t = \sigma_{n-2} \\ \chi_1 - \chi_0 = O(1/n) & X_t(i) = n, \eta_t = \sigma_n \\ w\chi_0 - \chi_0 = O(1/n) & X_t(i) = n, \eta_t = \sigma_{n-1} \\ w\chi_0 - \chi_0 = O(1/n) & X_t(i) = n, \eta_t = \sigma_{n-2} \\ 1 - \chi_1 = O(1/n) & X_t(i) = n-1, \eta_t = \sigma_n \\ 1 - \chi_1 = O(1/n) & X_t(i) = n-1, \eta_t = \sigma_{n-1} \\ w\chi_1 - \chi_1 = O(1/n) & X_t(i) = n-1, \eta_t = \sigma_{n-2} \end{cases}$$

Consequently,

$$|\Psi(X_{t+1}^{-1}, Y_{t+1}) - \Psi(X_t^{-1}, Y_t)| = O(1/n)$$

and we can take $R = O(1/n^2)$. The result follows by Proposition 4.2. $\square$

4.3 The Reversible Shuffle

Given a walk driven by a non-reversible probability measure q, common symmetric variants of the walk include $q^m \star q^{*m}$, $q^{*m} \star q^m$, and $\frac{1}{2}(q + q^*)$. In the case of top to bottom k_n, q^*_{n,k_n} can be interpreted as the walk that picks a card uniformly at random from the bottom k_n positions and moves it to the top. If $m \leq n - k_n$ then $q^m_{n,k_n} \star q^{*m}_{n,k_n}$ leaves the card in position $n - k_n + 1$ fixed, and $q^{*m}_{n,k_n} \star q^m_{n,k_n}$ leaves the first card fixed. Consequently, both of these reversible variants are irreducible if and only if $m > n - k_n$. In this section, we focus on the reversible walk $\frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$.

For reversible chains, path comparison is a useful technique for studying rates of convergence (see e.g. [DSC93a, DSC93b, DSC95a, DS91]). In particular, many of the arguments in this section rely on the notion of a flow to compare top to bottom shuffles with the well studied random transposition walk. Together with estimates on the least eigenvalue, this approach yields L^2 mixing time bounds.

To begin, consider a symmetric probability measure q on a finite group G and fix a symmetric set S that generates G and such that $q(s) > 0$ for $s \in S$. Define paths in the Cayley graph (G, S) to be sequences $\delta = (e, y_1, y_2, \ldots y_k)$ where e is the group identity and $y_i^{-1} y_{i+1} \in S$. Given such a path, define its length to be $|\delta| = k$, and for each $s \in S$, let

$$N(s, \delta) = \left| \left\{ i \in \{0, \ldots, k-1\} \mid y_i^{-1} y_{i+1} = s \right\} \right| .$$

That is, $N(s, \delta)$ is the number of times the generator s is used in the path δ. Furthermore, let $d_S(x, y)$ denote the graph distance on (G, S) between x and y.

Definition 4.1. Fix two symmetric probability measures $\tilde{q}, q$ on a finite group G and a symmetric set generating $S \subset \operatorname{supp}(q)$. A $(\tilde{q}, q)$-flow is a non-negative function η on the set of all paths $\mathcal{P}$ in the Cayley graph (G, S) such that

$$\sum_{\delta \in \mathcal{P}_y} \eta(\delta) = \tilde{q}(y)$$

where $\mathcal{P}_y$ is the set of all paths from the group identity e to y contained in $\mathcal{P}$.

4.3.1 The Least Eigenvalue

This section presents a lower bound on the smallest eigenvalue of the chain $\frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$. The proof relies on a geometric result that bounds the eigenvalues of symmetric chains by considering loops at the identity of odd length. (See [DSC95a] for details.) Together with comparison, Lemma 4.6 will be used to derive estimates on mixing time in Section 4.3.2.

The following definition of an odd flow is analogous to that of a flow, but restricted to paths of odd length.

Definition 4.2. Fix two symmetric probability measures $\tilde{q}, q$ on a finite group G and a symmetric set $S \subset \mathrm{supp}(q)$. A $(\tilde{q}, q)$-odd flow is a non-negative function η on the set of paths of odd length $\mathcal{O}$ in the Cayley graph (G, S) such that

$$\sum_{\delta \in \mathcal{O}_y} \eta(\delta) = \tilde{q}(y)$$

where $\mathcal{O}_y$ is the set of all paths of odd length from the group identity e to y contained in $\mathcal{O}$.

Note that we are not assuming that S generates G, i.e. the Cayley graph (G, S) need not be connected. However, the existence of a $(\tilde{q}, q)$-odd flow implies that for each y with $\tilde{q}(y) > 0$, there is at least one path from e to y in $\mathcal{O}$.

Proposition 4.3 ([DSC95a]). *Fix two symmetric probability measures $\tilde{q}, q$ on a group G and a symmetric generating set $S \subset supp(q)$. For any $(\tilde{q}, q)$-odd flow η,*

$$\beta_{min} \geq -1 + \frac{1 + \tilde{\beta}_{\min}}{A(\eta)}$$

where $\beta_{\min}$ and $\tilde{\beta}_{\min}$ are the smallest eigenvalues of q and $\tilde{q}$ respectively, and

$$A(\eta) = \max_{s \in S} \frac{1}{q(s)} \sum_{\delta \in \mathcal{O}} |\delta| N(s, \delta) \eta(\delta).$$

It is well known that a chain q is aperiodic if and only if the least eigenvalue satisfies $\beta_{\min} = -1$. As a trivial application of Proposition 4.3, by taking $S = \{e\}$ and $\tilde{q}(e) = 1$, we have $\beta_{\min} \geq -1 + 2q(e)$. When our chain puts no weight on the identity, the above result provides a way to capture more subtle effects of aperiodicity on the least eigenvalue.

Lemma 4.6. *Let $\beta_{\min}$ be the smallest eigenvalue of the symmetric chain $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$. Then*

$$\beta_{\min} \geq -1 + \frac{k_n - 1}{k_n(n-k_n+2)(n+1)}.$$

Proof. We will apply Proposition 4.3 with $\tilde{q}(e) = 1$ and $\tilde{q}(g) = 0$ otherwise. In this case, $\tilde{\beta}_{\min} = 1$. Let $S = \operatorname{supp}(\tilde{q}_{n,k_n})$. For l odd and such that $n - k_n + 1 \leq l \leq n$, define paths

$$\delta_l^{\pm 1} = (e, \sigma_l, \sigma_l^2, \ldots, \sigma_l^l)^{\pm 1}$$

and set $\mathcal{O} = \{\delta_l^{\pm 1} \mid l \text{ odd},\ n - k_n + 1 \leq l \leq n\}$. Let,

$$\begin{aligned}
\eta(\delta_l^{\pm 1}) &\equiv \frac{1}{2\sum_{\substack{n-k_n+1 \leq m \leq n \\ m \ odd}} \frac{1}{m^2}} \cdot \frac{1}{|\delta_l^{\pm 1}|^2} \\
&\leq \int_{n-k_n+2}^{n+1} \frac{1}{x^2} \cdot \frac{1}{l^2} \\
&= \frac{(n-k_n+2)(n+1)}{k_n - 1} \cdot \frac{1}{l^2}
\end{aligned}$$

and $\eta(\delta) = 0$ otherwise. Then,

$$\begin{aligned}
A(\eta) &\leq \frac{2k_n(n-k_n+2)(n+1)}{k_n-1} \max_{s \in S} \sum_{\delta \in \mathcal{O}} \frac{N(s,\delta)}{|\delta|} \\
&= \frac{2k_n(n-k_n+2)(n+1)}{k_n-1}.
\end{aligned}$$

The result follows from Proposition 4.3. □

Proposition 4.3 gives the best results when we can use short paths. In the case of $\frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$, for paths δ with $|\delta| \leq \lfloor \frac{n-k_n}{2} \rfloor$, the card originally in position $\lfloor \frac{n-k_n}{2} \rfloor + 1$ moves distance ± 1 at each step along the path. Consequently, the shortest loops at the identity with odd length have length $\approx n - k_n$.

4.3.2 Bounds on Mixing Times

The following lemma gives a lower bound on the mixing time of $\frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$ for k_n sufficiently small by looking at the motion of an individual particle.

Lemma 4.7. *Let* $\tilde{q}_{n.k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$ *with* $k_n \le cn$, $0 < c < 1$. *Then there is a constant* $N(c)$ *such that for* $n \ge N$, *and* $l \le \frac{c(1-c)^2n^2}{12}$,

$$\|\tilde{q}^l_{n,k_n} - \pi\|_{TV} \ge \frac{c}{2}.$$

In particular, there are constants $A(c) > 0$ *such that the total variation mixing time satisfies*

$$T_1 \ge An^2.$$

Proof. Note that the card originally in position $\lfloor \frac{(1-c)n}{2} \rfloor + 1$ performs a simple random walk on $\{1, \ldots, \lfloor (1-c)n \rfloor\}$ before hitting any of the bottom $\lfloor cn \rfloor$ positions. Call this card σ_a and define the event

$$A = \{\sigma \mid n - \lfloor cn \rfloor < \sigma^{-1}(\sigma_a) \le n\},$$

i.e. σ_a is in the bottom $\lfloor cn \rfloor$ positions. Then $\pi(A) \ge c - 1/n$. For $l = \lfloor \frac{c(1-c)^2n^2}{12} \rfloor$, let $X_1, \ldots, X_l$ be iid random variable with $P(X_i = \pm 1) = \frac{1}{2}$, and let $S_j = \sum_1^j X_i$. Then

$$\begin{aligned}
\tilde{q}^l_{n,k_n}(A) &\le P\left[\max_{1\le j\le l} |S_j| \ge \frac{(1-c)n}{2}\right] \\
&\le \frac{4l}{(1-c)^2n^2} \qquad \text{by Kolmogorov's maximal inequality} \\
&\le \frac{c}{3}.
\end{aligned}$$

Since $\|\tilde{q}^l_{n,k_n} - \pi\|_{TV} = \max_{A\subset S_n} |\tilde{q}^l_{n,k_n}(A) - \pi(A)|$, the result follows by taking n sufficiently large. The mixing time bound follows from the fact that for n sufficiently large,

$$c \le 2\|\tilde{q}^{l}_{n,k_n} - \pi\|_{TV} \le e^{-\lfloor l/T(S_n,\tilde{q}_{n,k_n})\rfloor}.$$

In particular,

$$T_1 \ge \frac{l}{1-\log c}.$$

□

Now we will derive an upper bound on the mixing time of $\frac{1}{2}(q_{n,k_n}+q^*_{n,k_n})$ with $n-k_n \le C$ independent of n. That is, the symmetric version of the walk that moves the top card uniformly at random to any but a finite number of the top positions. The proof is by comparison and is based on the following two results. For proofs of these results see e.g. [DSC95a, DSC93a, DSC93b].

Recall that on a finite group G equipped with a probability measure q, the Dirichlet form is given by

$$\mathcal{E}_q(f,f) = \frac{1}{2|G|}\sum_{x,y}|f(xy)-f(x)|^2 q(y).$$

Proposition 4.4. *Assume that $q, \tilde{q}$ are symmetric measures and $\mathcal{E}_{\tilde{q}} \le A\mathcal{E}_q$. Then,*

$$T_1(G,q) \le T_2(G,q) \preceq \max\left\{AT_2(G,\tilde{q}), A\log|G|, \frac{1}{-\log\beta_-}\right\}$$

where $\beta_- = \max\{0, -\beta_{\min}\}$.

Proposition 4.5. *For symmetric measures $q, \tilde{q}$ and any $(\tilde{q},q)$-flow, $\mathcal{E}_{\tilde{q}} \le A\mathcal{E}_q$ with*

$$A(\eta) = \max_{s\in S}\frac{1}{q(s)}\sum_{\delta\in\mathcal{P}}|\delta|N(s,\delta)\eta(\delta).$$

The proofs of the following two mixing time bounds are by comparison with the random transposition measure on S_n:

$$q_{RT,n}(g) \quad = \quad \begin{cases} 1/n & \text{if } g=e \\ 2/n^2 & \text{if } g=(i,j),\ i\neq j \\ 0 & \text{otherwise} \end{cases}$$

Lemma 4.8. *For $k_n \geq n - C$, there exist constants $B(C)$ such that $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$ has L^2 mixing time satisfying*

$$T_2 \leq Bn \log n.$$

Proof. Let $S = \{\sigma_l^{\pm 1}\}$. First we define paths $\delta_{i,j}$, $1 \leq i < j \leq n$ from e to (i, j) in the Cayley graph (S_n, S).

$$\delta_{i,j} = \begin{cases} \sigma_i^{-1}\sigma_j\sigma_{j-1}^{-1}\sigma_i & C+1 \leq i < j \leq n \\ (\sigma_n^{-1})^{C-i+1}\,\sigma_{C+1}^{-1}\,\sigma_{j+C-i+1}\,\sigma_{j+C-i}^{-1}\,\sigma_{C+1}\,\sigma_n^{C-i+1} & 1 \leq i \leq C, \\ & i < j \leq n - C \\ (\sigma_{n-C}^{-1})^{C-i+1}\,\sigma_{C+1}^{-1}\,\sigma_j\,\sigma_{j-1}^{-1}\,\sigma_{C+1}\,\sigma_{n-C}^{C-i+1} & 1 \leq i \leq C, \\ & j > n - C \end{cases}$$

Define a $(q_{RT,n}, \tilde{q}_{n,k_n})$ flow by $\eta(\delta_{i,j}) = \frac{1}{n^2}$. For $i \leq C$, $|\delta_{i,j}| \leq 2(C-2)$. And each $s \in S$ is used in at most n paths $\delta_{i,j}$ with $i > C$. Consequently,

$$A(\eta) \leq 8[C(C+2)^2 + 1].$$

Since $T_1(S_n, q_{RT}) \sim \frac{n}{2}\log n$ (see [DS81] for details), the result follows by applying Lemma 4.6 and Proposition 4.4 together with Proposition 4.5 □

The following lemma bounds the mixing time of $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$ for arbitrary k_n. The proof is by comparison with the random transposition measure, but while the flow defined in Lemma 4.8 used only one path for each transposition, here for most transpositions we define $k-1$ paths.

Lemma 4.9. *There exists a constant A such that the L^2 mixing time for $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$ satisfies*

$$T_2 \leq An^3 \log n.$$

Proof. Let $S = \{\sigma_l^{\pm 1}\}$. The proof is by comparison with the random transposition measure q_{RT}. For $j > i > n-k$, define the path

$$\delta_{i,j} = \sigma_i^{-1}\sigma_j\sigma_{j-1}^{-1}\sigma_i.$$

For $i < n-k$, we define $k-1$ distinct paths $\delta_{i,j}^l$ with $n-k < l < n$. For $j > l$, let

$$\delta_{i,j}^l \equiv (\sigma_l^{-1})^{l-i}\,\delta_{l,j}\,\sigma_l^{l-i}$$

and for $i < j \leq l$, let

$$\delta_{i,j}^l \equiv (\sigma_l^{-1})^{l-j}\,\delta_{l,l+1}(\sigma_l)^{j-i}\,\delta_{l,l+1}\,\sigma_l^{j-i}\,\delta_{l,l+1}\,\sigma_l^{l-j}.$$

So $|\delta_{i,j}^l| \leq 2n + 12 \leq 3n$. Define a $(q_{RT}, \tilde{q}_{n,k_n})$-flow by $\eta(\delta_{i,j}) = \frac{1}{n^2}$ and $\eta(\delta_{i,j}^l) = \frac{1}{(k-1)n^2}$. Then,

$$\begin{aligned} A(\eta) &\leq \frac{6}{n}\max_s \sum_{\delta_{i,j}^l} N(s,\delta_{i,j}^l) + \frac{8k}{n^2}\max_s \sum_{\delta_{i,j}} N(s,\delta_{i,j}) \\ &\leq 18n^2 + \frac{8k^2}{n^2}. \end{aligned}$$

Since $T_1(S_n, q_{RT}) \sim \frac{n}{2}\log n$ (see [DS81] for details), the result follows by applying Lemma 4.6, Proposition 4.4 and Proposition 4.5. □

The following lemma shows the difficulty in applying path comparison via Proposition 4.5 to bound mixing time.

Lemma 4.10. *Consider a $(\tilde{q}, q)$-flow η on (G, S). For*

$$A(\eta) = \max_{s\in S}\frac{1}{q(s)}\sum_{\delta\in P}|\delta|N(s,\delta)\eta(\delta)$$

we have the lower bound

$$A(\eta) \geq \sum_{g\in G} d_S^2(e,g)\tilde{q}(g).$$

In particular, for $X \subset G$, $A(\eta) \geq d_S^2(e,X)\tilde{q}(X)$.

Proof. By averaging over s,

$$\begin{aligned} A(\eta) &\geq \sum_{s,\delta} |\delta| N(s,\delta)\eta(\delta) \\ &= \sum_{\delta} |\delta|^2 \eta(\delta) \\ &\geq \sum_{g} d_S^2(e,g)\tilde{q}(g). \end{aligned}$$

□

Observe that we can always choose a $(\tilde{q}, q)$-flow η such that

$$A(\eta) \leq \left(\max_{s\in S} \frac{1}{q(s)}\right) \sum_{g\in G} d_S^2(e,g)\tilde{q}(g).$$

Lemma 4.10 shows that upper bounds on mixing time that we derive in this section are the best one can do using comparison with the random transposition walk.

Consider a symmetrized variant of the Rudvalis shuffle driven by the measure $\dot{q}_n$ which is uniform on the generating set $\{\sigma_n, \sigma_n^{-1}, (1,n), id\}$. This walk was analyzed in [Wil03b] and an $O(n^3 \log n)$ lower bound was derived for the total variation mixing time (see e.g. [SC04a] for a matching upper bound). Here we use comparison to extend this result to lower bounds for symmetrized top to bottom walks.

Lemma 4.11. *For $0 < c < 1$, and $k_n \leq cn$, there exist constants C, $N(c)$ such that for $\tilde{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + q^*_{n,k_n})$, and $n \geq N(c)$, the L^2 mixing time satisfies*

$$T_2 \geq \frac{Cn^3}{k_n^2} \log n.$$

Proof. Let $S = \{\sigma_n^{\pm 1}, \tau\}$, where $\tau = (1,n)$ and observe that

$$\begin{aligned} \sigma_l &= \sigma_n \cdot (\sigma_{n-1}^{-1})^{n-l} \cdot \sigma_n^{n-l} \\ &= \sigma_n \cdot (\sigma_n^{-1} \cdot \tau)^{n-l} \cdot \sigma_n^{n-l} \end{aligned}$$

For $n - k_n < l \leq n$, define paths $\delta_{\sigma_l^{\pm 1}}$ in the Cayley graph (S_n, S) by the above, and a corresponding simple $(\tilde{q}_{n,k_n}, \dot{q}_n)$-flow η. Then

$$\begin{aligned} A(\eta) &\leq \frac{4}{k_n} \sum_{n-k_n<l\leq n} |\delta_{\sigma_l}|^2 \\ &= \frac{4}{k_n} \sum_{n-k_n<l\leq n} [3(n-l)+1]^2 \\ &\leq Bk_n^2 \end{aligned}$$

for some universal constant B. By Proposition 4.5, $\mathcal{E}_{\tilde{q}_{n,k_n}} \leq Bk_n^2 \mathcal{E}_{\dot{q}_n}$. By Proposition 4.4 together with the lower bound on the mixing time for $\dot{q}_n$ given in [Wil03b], we have

$$n^3 \log n \preceq \max\left\{AT_2(G, \tilde{q}), A\log|G|, \frac{1}{-\log\beta_-}\right\}.$$

By Lemma 4.6, $-1/\log\beta_- = O(n^2)$, and so either $AT_2(G,\tilde{q})$ or $A\log|G|$ is bounded below by $n^3\log n$. By Lemma 4.7 $n^2 \preceq T_2(G,\tilde{q})$, and so $AT_2(G,\tilde{q}) > A\log|G|$. Consequently, for n sufficiently large

$$n^3 \log n \leq AT_2(G,\tilde{q})$$

and the result follows. $\square$

4.4 The Lazy Shuffle

We show that our estimates on the mixing times for $\tilde{q}_{n,k_n}$ and q_{n,k_n} yield bounds for the lazy top to bottom shuffles. In order to transfer mixing time results for the reversible walk $\tilde{q}_{n,k_n}$ to the present case of

$$\hat{q}_{n,k_n} = \frac{1}{2}(q_{n,k_n} + \delta_e)$$

we recall the following result.

Proposition 4.6 ([DSC95a]). *Let q be a probability measure on G and set $q_* = q \star q^*$. Then*

$$T(G,q) \leq T_2(G,q) \leq 2T_2(G,q_*).$$

More generally, if $q_v = q^v \star q^{*v}$, then $T_2(G,q) \leq 2vT_2(G,q_v)$. Finally, $q^{*v} \star q^v$ can be used instead of q_v.

Lemma 4.12. *For $k_n \geq n-C$, there exist constants $B(C)$ such that the L^2 mixing time for $\hat{q}_{n,k_n}$ satisfies*

$$T_2 \leq Bn \log n.$$

For arbitrary k_n, there is a constant A such that

$$T_2 \leq An^3 \log n.$$

Proof. By Proposition 4.6, it is sufficient to prove the bounds for

$$p_{n,k_n} = \hat{q}^*_{n,k_n} \star \hat{q}_{n,k_n}.$$

Observe that,

$$\begin{aligned} p_{n,k_n} &= \frac{1}{2}(q^*_{n,k_n} + \delta_e) \star \frac{1}{2}(q_{n,k_n} + \delta_e) \\ &= \frac{1}{2}\left[\tilde{q}_{n,k_n} + \frac{1}{2}(q^*_{n,k_n} \star q_{n,k_n} + \delta_e)\right] \\ &\geq \frac{1}{2}\tilde{q}_{n,k_n}. \end{aligned}$$

Consequently, $\mathcal{E}_{\tilde{q}_{n,k_n}}(f,f) \leq 2\mathcal{E}_{p_{n,k_n}}(f,f)$. Note that p_{n,k_n} is a positive operator and consequently has non-negative eigenvalues. The result then follows from Proposition 4.4 together with the L^2 mixing time bounds for $\tilde{q}_{n,k_n}$ derived in Section 4.3.2. □

To transfer total variation mixing time results for q_{n,k_n} to the lazy top to bottom shuffle, we make the following elementary observation.

Definition 4.3. Let q drive a walk on G. Then for $p \in (0, 1)$ the kernel of the associated p-lazy walk is given by

$$\hat{q}_p = pq + (1-p)\delta_e.$$

Lemma 4.13. *Let q drive a walk on G with stationary distribution π, and fix $p, \epsilon \in (0, 1)$. Then there exist constants $C(p, \epsilon)$ such that mixing times for q and the associated p-lazy walk $\hat{q}_p$ satisfy*

$$T_1(G, \hat{q}_p) \le \max\left[\frac{2+\epsilon}{p} T_1(G, q), C\right].$$

Specifically, we can take $C = 80/(p\epsilon^2)$.

Proof. Let S_m be a binomial random variable with parameters m and p. Then

$$\begin{aligned}
\|\hat{q}_p^m - \pi\|_{TV} &= \frac{1}{2} \sum_{g \in G} |\hat{q}_p^m(g) - \pi(g)| \\
&= \frac{1}{2} \sum_{g \in G} \left| \sum_k P(S_m = k)(q^k(g) - \pi(g)) \right| \\
&\le \sum_k P(S_m = k) \cdot \|q^k - \pi\|_{TV} \\
&\le P(S_m \le 2T_1(G, q)) + \frac{1}{2e^2}.
\end{aligned}$$

Taking $\bar{m} \ge \frac{2+\epsilon}{p} T(G, q)$, by Chebyshev's inequality,

$$\begin{aligned}
P(S_{\bar{m}} \le 2T_1(G, q)) &\le P\left(|S_{\bar{m}} - ES_{\bar{m}}| \ge \left(1 - \frac{2}{2+\epsilon}\right) ES_{\bar{m}}\right) \\
&\le \frac{1-p}{\bar{m}p(1 - \frac{2}{2+\epsilon})^2}.
\end{aligned}$$

And consequently,

$$\begin{aligned}
\|\hat{q}_p^{\bar{m}} - \pi\|_{TV} &\le \frac{1-p}{\bar{m}p(1 - \frac{2}{2+\epsilon})^2} + \frac{1}{2e^2} \\
&\le \frac{1}{2e}
\end{aligned}$$

for $\bar{m} \ge 80/(p\epsilon^2)$. $\square$

Now we can transfer the mixing time results for q_{n,k_n} to $\hat{q}_{n,k_n}$.

Corollary 4.1. *For $k_n \geq n - \sqrt{(n \log n)/2}$, there exists a constant C such that the L^1 mixing time for $\hat{q}_{n,k_n}$ satisfies*

$$T_1 \leq Cn \log n.$$

For $c \in (0,1)$ and $k_n \geq cn$, there exist constants $A(c)$ such that

$$T_1 \leq An^2 \log^2 n.$$

Remark 4.3. For $k_n \geq n - \sqrt{(n \log n)/2}$, instead of using Lemma 4.13, we can adapt the coupling of Lemma 4.2 to show $T(S_n, \hat{q}_{n,k_n}) \sim 2n \log n$. The coupling (X_1^m, X_2^m) of q_{n,k_n} lifts to a coupling

$$(\tilde{X}_1^m, \tilde{X}_2^m) = (X_1^{S_m}, X_2^{S_m})$$

of $\hat{q}_{n,k_n}$ where S_m is an independent binomial$(1/2, m)$ random variable. Then, if T is the coupling time for (X_1^m, X_2^m)

$$P(\tilde{X}_1^m \neq \tilde{X}_2^m) \leq P\left(S_m \leq \left(1 + \frac{\epsilon}{2}\right) n \log n\right) + P\left(T > \left(1 + \frac{\epsilon}{2}\right) n \log n\right).$$

For $m = 2(1+\epsilon)n \log n$ the first term goes to 0 by Chebyshev's inequality, and the second term goes to 0 by the cut-off shown in Lemma 4.2.

The lower bound is also analogous to that given in Lemma 4.2, where we now make the observation that

$$P(\hat{L}_j > m) \geq P\left(L_j > \left(1 - \frac{\epsilon}{2}\right) n \log n\right) \cdot P\left(S_m \leq \left(1 - \frac{\epsilon}{2}\right) n \log n\right).$$

So for $k_n \geq n - \sqrt{(n \log n)/2}$ and $\epsilon \in (0,1)$

$$\lim_{n \to \infty} \|\hat{q}_{n,k_n}^{(1-\epsilon)2n \log n} - \pi\|_{TV} = 1$$

and

$$\lim_{n \to \infty} \|\hat{q}_{n,k_n}^{(1+\epsilon)2n \log n} - \pi\|_{TV} = 0.$$

Finally, transferring the lower bounds for $k_n = 2, 3$ which were derived using Wilson's lemma also requires only a simple argument. Let $\{\eta_i\}$ be iid Bernoulli random variables with $p = 1/2$, and let $N_t = \sum_{i=1}^{t} \eta_i$. Then if X_t is the top to bottom k_n process, the lazy top to bottom k_n process is given by $\tilde{X}_t = X_{N_t}$. Using the notation of Proposition 4.2, if (X_t, Y_t) is a lifting of X_t, then $(\tilde{X}_t, \tilde{Y}_t) = (X_{N_t}, Y_{N_t})$ is a lifting of $\tilde{X}_t$. It is not hard to check that the assumptions of the theorem are met with $\tilde{\Psi} = \Psi$, $\tilde{\lambda} = 1/2 + 1/2\lambda$, $\tilde{R} = R/2$, and $\tilde{\gamma} = \gamma/2$. Then

$$\frac{\log \tilde{\Psi}_{\max} + \frac{1}{2} \log \frac{\tilde{\gamma}\epsilon}{4\tilde{R}}}{-\log(1-\tilde{\gamma})} = \frac{\log \frac{\Psi_{\max}}{2} + \frac{1}{2} \log \frac{\gamma\epsilon}{4R}}{-\log(1-\gamma/2)}.$$

Using the estimates in Lemma 4.5 and [Wil03b], we have the following lower bounds.

Corollary 4.2. *For $k_n = 2, 3$ and $\epsilon > 0$, there exist constants $C(\epsilon) > 0$ such that the lazy top to bottom shuffle satisfies*

$$\|\hat{q}^m_{n,k_n} - \pi\|_{TV} \geq 1 - \epsilon$$

for $m \leq Cn^3 \log n$.

BIBLIOGRAPHY

[ABC+00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de sobolev logarithmiques*, Société Mathématique de France, 2000.

[AF] D. Aldous and J. Fill, *Reversible markov chains and random walks on graphs*, Monograph in preparation available on the web at http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[AL00] Cécile Ané and Michel Ledoux, *On logarithmic sobolev ineqalities for continuous time random walks on graphs*, Prob. Theory Relat. Fields **116** (2000), 573–602.

[Ald82] D. Aldous, *Random walks on finite groups and rapidly mixing markov chains*, Séminaire de Probabilités XVII (J. Azéma and M. Yor, eds.), Lecture Notes in Mathematics, vol. 986, Springer, 1981/82.

[Alo87] N. Alon, *Eigenvalues and expanders*, Combinatorica **5** (1987), 83–96.

[BCG01] Martin Barlow, Thierry Coulhon, and Alexander Grigor'yan, *Manifolds and graphs with slow heat kernel decay*, Invent. Math. **144** (2001), 609–649.

[BHT] S.G. Bobkov, C. Houdré, and P. Tetali, *The subgaussian constant and concentration inequalities*, A preprint is available on the web at http://www.math.gatech.edu/~houdre/.

[BL98] S.G. Bobkov and M. Ledoux, *On modified logarithmic sobolev inequalities for bernoulli and poisson measures*, J. Funct. Anal. **156** (1998), 347–365.

[BL00] ______, *From brunn-minkowski to brascamp-lieb and to logarithmic sobolev inequalities*, Geom. funct. anal. **10** (2000), 1028–1052.

[Bol80] B. Bollobas, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), 311–316.

[BT] Sergey G. Bobkov and Prasad Tetali, *Modified logarithmic sobolev inequalities in discrete settings*, Preprint available on the web at http://www.math.gatech.edu/~tetali/.

[BT03] ______, *Modified log-sobolev inequalities in discrete settings*, Proc. of the ACM STOC 2003, 2003, pp. 287–296.

[CGP01] T. Coulhon, A. Grigor'yan, and C. Pittet, *A geometric approach to on-diagonal heat kernel lower bound on groups*, Ann. Inst. Fourier **51** (2001), 1763–1827.

[CKS87] E. Carlen, S. Kusuoka, and D. Stroock, *Upper bounds for symmetric markov transition functions*, Ann. Inst. H. Poincaré Probab. Statist. **23** (1987), 245–287.

[Cou96] T. Coulhon, *ultracontractivity and nash type inequalities*, J. Funct. Anal. **141** (1996), 510–539.

[CSC90a] Thierry Coulhon and Laurent Saloff-Coste, *Marches aléatoires non symétriques sur les groupes unimodulaires*, C. R. Acad. Sci. Paris Sér. I Math. **310** (1990), no. 8, 627–630.

[CSC90b] ———, *Puissances d'un op'erateur régularisant*, Ann. Inst. H. Poincaré Probab. Statist. **26** (1990), no. 3, 419–436.

[Dav90] E.B. Davies, *Heat kernels and spectral theory*, Cambridge University Press, 1990.

[DFK91] M. Dyer, A. Frieze, and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Journal of the ACM **38** (1991), 1–17.

[DFP92] Persi Diaconis, Jim Fill, and Jim Pitman, *Analysis of top to random shuffles*, Combinatorics, Probability Computing **1** (1992), 135–155.

[Dia88] Persi Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics, 1988.

[Dia92] ———, *Finite fourier methods: Access to tools*, Probabilistic Combinatorics and Its Applications (Béla Bollobás, ed.), Proceedings of Symposia in Applied Mathematics, vol. 44, American Mathematical Society, 1992.

[DS81] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transposition*, Z. Wahr. Verw. Geb. **57** (1981), 159–179.

[DS87] ———, *Time to reach stationarity in the bernoulli-laplace diffusion model*, SIAM J. Math. Anal. **18** (1987), 208–218.

[DS91] P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues for markov chains*, Ann. Appl. Prob. **1** (1991), 36–61.

[DSC93a] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible markov chains*, Ann. Appl. Prob. **3** (1993), 696–730.

[DSC93b] Persi Diaconis and Laurent Saloff-Coste, *Comparison techniques for random walk on finite groups*, Ann. Prob. **21** (1993), 2131–2156.

[DSC94] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk on finite groups*, Geom. and Funct. Anal. **4** (1994), 1–34.

[DSC95a] ______, *Random walks on finite groups: a survey of analytic techniques*, Probability measures on groups and related structures, Oberwolfach, World Sci. Publishing, 1995.

[DSC95b] Persi Diaconis and Laurent Saloff-Coste, *An application of harnack inequalities to random walk on nilpotent quotients*, Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993), J. Fourier Anal. Appl, Special Issue, 1995, pp. 189–207.

[DSC96a] P. Diaconis and L. Saloff-Coste, *Logarithmic sobolev inequalities for finite markov chains*, Ann. Appl. Prob. **6** (1996), 695–750.

[DSC96b] ______, *Nash inequalities for finite markov chains*, J. Theoret. Probab. **9** (1996), no. 2, 459–510.

[Fel68] W. Feller, *An introduction to probability theory and its applications*, vol. I, Wiley, 1968.

[FOW85] L. Flatto, A.M. Odlyzko, and D.B. Wales, *Random shuffles and group representations*, Ann. Prob. **13** (1985), 154–178.

[GMT] Sharad Goel, Ravi Montenegro, and Prasad Tetali, *Mixing time bounds via the spectral profile*, Elec. J. of Prob., submitted.

[Goe] Sharad Goel, *Analysis of top to bottom-k shuffles*, Annals of Applied Prob., to appear.

[Goe04] ______, *Modified logarithmic sobolev inequalities for some models of random walk*, Stoc. Proc. Appl. **114** (2004), 51–79.

[GQ03] Fuqing Gao and Jeremy Quastel, *Expnential decay of entropy in the random transposition and bernoulli-laplace models*, Ann. Appl. Prob. **4** (2003), 1591–1600.

[Gri94] A. Grigor'yan, *Heat kernel upper bounds on a complete non-compact manifold*, Revista Matemática Iberoamericana **10** (1994), 395–452.

[Gro75] L. Gross, *Logarithmic sobolev inequalities*, Amer. J. Math. **97** (1975), 1061–1083.

[Gro93] ______, *Logarithmic sobolev inequalities and contractivity properties of semigroups*, Lecture Notes in Mathematics, vol. 1563, Springer, 1993.

[Hil90] M. Hildebrand, *Rates of convergence of some random processes on finite groups*, Ph.D. thesis, Harvard University, 1990.

[HY] Y. Higuchi and N. Yoshida, *Analytic conditions and phase transition for ising models*, Unpublished lecture notes (in Japanese).

[Jon] Johan Jonasson, *Biased random-to-top shuffling*, preprint.

[JS93] M. Jerrum and A. Sinclair, *Polynomial-time approximation algorithms for the ising model*, SIAM Journal on computing **22** (1993), 1087–1116.

[JS96] Mark Jerrum and Alistair Sinclair, *The markov chain monte carlo method: an approach to approximate counting and integration*, Approximation Algorithms for NP-hard Problems (D.S. Hochbaum, ed.), PWS Publishing, Boston, 1996.

[Led01] Michel Ledoux, *The concentration of measure phenomenon*, American Mathematical Society, 2001.

[LK99] L. Lovász and R. Kannan, *Faster mixing via average conductance*, Annual ACM Symposium on theory of computing (Atlanta, GA, 1999), ACM, 1999, pp. 282–287.

[LO00] R. Latala and K. Oleszkiewicz, *Between sobolev and poincaré*, Geometric Aspects of Functional Analysis, Lect. Notes in Math., vol. 1745, 2000, pp. 147–168.

[LS88] G.F. Lawler and A.D. Sokal, *Bounds on the L^2 spectrum for markov chains and markov processes: a generalization of cheeger's inequality*, Trans. Amer. Math. Soc. **309** (1988), 557–580.

[LY98] Tzong-Yow Lee and Horng-Tzer Yau, *Logarithmic sobolev inequalities for some models of random walks*, Ann. Prob. **26** (1998), 1855–1873.

[MP] Ben Morris and Yuval Peres, *Evolving sets, mixing and heat kernel bounds*, Preprint.

[MS93] N. Madras and G. Slade, *The self-avoiding walk*, Birkhäuser, Boston, 1993.

[Nas58] J. Nash, *Continuity of solutions of parabolic and elliptic equations*, Amer. J. Math **80** (1958), 931–954.

[PSC] Ch. Pittet and L. Saloff-Coste, *Isoperimetry, volume growth and random walks*, Monograph in preparation.

[Rob01] Cyril Roberto, *Inégalités de hardy et de sobolev logarithmiques*, Ph.D. thesis, Université Paul Sabatier, Toulouse III, June 2001.

[Rou99] Sandrine Roussel, *Marches aléatoires sur le groupe symétrique*, Ph.D. thesis, Université Paul Sabatier, Toulouse III, June 1999.

[Rou00] ———, *Phénomène de cutoff pour certaines marches aléatoires sur le groupe symétrique*, Colloq. Math. **86** (2000), 111–135.

[SC94] L. Saloff-Coste, *Precise estimates on the rate at which certain diffusions tend to equilibrium*, Math.Z. **217** (1994), 641–677.

[SC96] Laurent Saloff-Coste, *Lectures on finite markov chains*, Lectures on Probability Theory and Statistics (P. Bernard, ed.), Lecture Notes in Mathematics, vol. 1665, Ecole d'Eté de Probabiltés de Saint-Flour XXVI, Springer, 1996.

[SC04a] ______, *Random walks on finite groups*, Probability on Discrete Structures (H. Kesten, ed.), Encyclopaedia of Mathematical Sciences, vol. 110, Springer, 2004.

[SC04b] ______, *Total variation lower bounds for finite markov chains: Wilson's lemma*, Random Walks and Geometry: Proceedings of a Workshop at the Erwin Schrödinger Institute, Vienna, June 18-July 13, 2001 (V. Kaimanovich, ed.), Walter De Gruyter Inc, 2004.

[SJ89] A.J. Sinclair and M. R. Jerrum, *Approximate counting, uniform generation and rapidly mixing markov chains*, Information and Computation **82** (1989), 93–133.

[Wil03a] Elizabeth Wilmer, *A local limit theorem for a family of non-reversible markov chains*, Journal of Theoretical Probability **16** (2003), 751–770.

[Wil03b] David B. Wilson, *Mixing time of the rudvalis shuffle*, Electron. Comm. Probab. **8** (2003), 77–85.

[Wil04] ______, *Mixing times of lozenge tiling and card shuffling markov chains*, Ann. Appl. Probab. **1** (2004), 274–325.

[Wu00] Liming Wu, *A new modified logarithmic sobolev inequality for poisson point processes and several applications*, Probab. Theory Relat. Fields **118** (2000), 427–438.